

Db 1354 TATGCCCTTCTAGTGGCTATTTTGGCGGTGTCATGGTTCGACTTATGCTTGTTCATCAG 1413
Qy 466 ProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGly 485
Db 1414 CTTGTTGCTGTGTTCTCGCATTTGGCTTCTCCAAACTTCTCGAGCGGTATATT--- 1470
Qy 486 AspAspMetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArg 505
Db 1471 -----GACCCCGTCATCCCGAAGCGACGAGGAAGTCGCGAG 1509
Qy 506 AsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGlu 525
Db 1510 TCTCAGACGCGGTGTCTCAAGTCC---AAGCGCAAGAGATGCGCGCTGCCACGCC 1566
Qy 526 -----LysThrGluGluGlyLeu--- 531
Db 1567 AATAAGACGGGTCTCTTTACAGGTATTTTGGCGGCAAGTCTGTCTCCGCGCATTTT 1626
Qy 532 GlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAla 551
Db 1627 GGTCTCGACACTCGATTGCTGTGGTTCCATTCTCTGTCTCTCTCTCTCTCTCTCT 1686
Qy 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerProSerValValLeuAlaSer 571
Db 1687 CTTCACTGCAATATGTACTTCAACAGCGTATTCTTCGCTTCACTGTGTACTTGTCAICG 1746
Qy 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeu 591
Db 1747 CGAAACCGGATGTTAGCCAAATAATCATGATGTTCCGAGAGCTTACTACTGGATT 1806
Qy 592 ArgGluAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
Db 1807 CGCAAAACACCCCGCAAGACAGCGTCATCATGCTCTGCTGGTGGATTACGGCTACCAGATC 1866
Qy 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631
Db 1867 GCTGGTATGCTGATCGCCCCACCTTGTGTGATAACAATACCTGGAATAACCCACAT 1926
Qy 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
Db 1927 GCCACAGTTGGTAAGCCATGGCTTCCACGAGATGTCGATATCTTCTTGGGAAG 1986
Qy 652 LeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAsp 671
Db 1987 CATGATGTCTGATACCTCTTGTGATCTTGGGGCTTATGGGTACTCTCTGTCAGCAT 2046
Qy 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691
Db 2047 ATCAACAGTTTGTGGATGTTAGATCTCACAGGTGATGGCTGACGAGGTGAG 2106
Qy 692 GluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThr 711
Db 2107 GAAGTCAACTACTTTACTCAAAAGAGGGAGTATGCTGTCGATCAGAGGCCACCCCTACT 2166
Qy 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTrpArgPheGlyGluMetGlnLeuAsp 731
Db 2167 ATGAAGACTCTCTCATGTACAAATGCTTACTACCGTCTCCCGAGCTT----- 2217
Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIle 751
Db 2218 TATGGTGGACACCCGCTCAAGACAGGGTTCGAGGCCAAATATCCCTCTTAACAGTGT 2277
Qy 752 LysPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLys 771
Db 2278 ACTCTTGATACTTTGACCAAGCGTTTCATCCGAAATGATGATCGTCAGGATCTACAAG 2337
Qy 772 ValLysAlaProAsp-----AsnArgGluThrLeuAspHisLys 784
Db 2338 GTCAAGAAGAGATCCCATTTGACGA-----GACCAACAG 2373

RESULT 14

US-10-128-714-7139

; Sequence 7139, Application US/10128714

Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wendi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Lemieux, Sebastien M
APPLICANT: Ershkin, Alexey M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7139
LENGTH: 2332
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-7139

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 9,46e-216 | Length: | 2232 |
| Score: | 2179.50 | Matches: | 416 |
| Percent Similarity: | 71.35% | Conservative: | 117 |
| Best Local Similarity: | 55.69% | Mismatches: | 177 |
| Query Match: | 49.52% | Indels: | 37 |
| DB: | 15 | Gaps: | 8 |

US-10-028-384-2 (1-826) x US-10-128-714-7139 (1-2332)

| | | | |
|----|-----|--|-----|
| Qy | 69 | GlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerArg | 88 |
| Db | 55 | CGCGAGCTTCTGGGATTATTATCTGTGTACCATTCGACGACGAGTTCGAGTCGA | 114 |
| Qy | 89 | LeuPheAlaValIleArgPheGluSerIleIleHisGluPheAspPheAspPheAspTyr | 108 |
| Db | 115 | CTCTTACGCTTATCGGCTTCGAGAGTATCATCCAGAGTTGACCGGTTCACATTC | 174 |
| Qy | 109 | ArgSerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGlu | 128 |
| Db | 175 | CGAGCAACAAATACCTAGTACAGAAATGGTTCTATAGCTTTTGGGATTTGGTATGAC | 234 |
| Qy | 129 | ArgAlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIle | 148 |
| Db | 235 | GGACATGGCATCTCTGGGACGTGTCACCGGTGGCAGCTTATATCCCGTCTCATGGTG | 294 |
| Qy | 149 | ThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAsp | 168 |
| Db | 295 | ACGAGCGCGTGTATCTACCATATCTTGGGATCTTCTTACTATCCCGTCGATATTCGCAAC | 354 |
| Qy | 169 | ValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeu | 188 |
| Db | 355 | ATCTGCTCTCTACTGGCGCAGGATCTCCGGCTGACTGCAATGGCAATGCTACTTGGTG | 414 |
| Qy | 189 | ThrArgGluLeuTrp---AsnGlnGlyAlaGlyLeuAlaAlaCysPheIleAlaIle | 207 |
| Db | 415 | ACATCCGAGATGCTCTCTCGCCATCTGAGGCTCTTCTTGGCAGCACTTTCATGGGAATC | 474 |
| Qy | 208 | ValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIle | 227 |
| Db | 475 | GCCTCTGTTACATCTCCCGATCATGTTGTTGGAAGCTACGATAACGAAGCGATGCGATC | 534 |

| | | | | | | | |
|---|----------|--|------|----|------|--|------|
| Qy | 764 | HistHrLeuValArgIleTyrLys | 771 | Qy | 270 | IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg | 289 |
| Db | 2191 | CATTGGCTGGTCAGGATATACAG | 2214 | Db | 781 | ATCAACTTAATTCCTCCACGTCCTGCTGATGCTCACAGGCCCTTTCTCTCACCGG | 840 |
| RESULT 12 | | | | | | | |
| US-10-417-375-97 | | | | | | | |
| ; Sequence 97, Application US/10417375 | | | | | | | |
| ; Publication No. US20040219528A1 | | | | | | | |
| ; GENERAL INFORMATION: | | | | | | | |
| ; APPLICANT: David W. Morris | | | | | | | |
| ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer | | | | | | | |
| ; FILE REFERENCE: 529452001600 | | | | | | | |
| ; CURRENT APPLICATION NUMBER: US/10/417,375 | | | | | | | |
| ; NUMBER OF SEQ ID NOS: 176 | | | | | | | |
| ; SOFTWARE: Fast-Seq for Windows Version 4.0 | | | | | | | |
| ; SEQ ID NO 97 | | | | | | | |
| ; LENGTH: 5827 | | | | | | | |
| ; TYPE: DNA | | | | | | | |
| ; ORGANISM: Homo sapiens | | | | | | | |
| US-10-417-375-97 | | | | | | | |
| Alignment Scores: | | | | | | | |
| Pred. No.: | 1.8e-221 | Length: | 5827 | Qy | 70 | SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu | 89 |
| Score: | 2241.00 | Matches: | 424 | Db | 181 | ACACTTTGAAGCTTCTCATCTCTGTCATGGCTGCTGCTATATCTCTCCACTGCTG | 240 |
| Percent Similarity: | 74.44% | Conservative: | 103 | Qy | 90 | PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheSerIleArg | 109 |
| Best Local Similarity: | 59.89% | Mismatches: | 145 | Db | 241 | TTTGCTGCTGAGATTGAAGTGTATCCATGAGTTGATCCCGTACTTTAATATCGG | 300 |
| Query Match: | 50.92% | Indels: | 36 | Qy | 110 | SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg | 129 |
| DB: | 18 | Gaps: | 7 | Db | 301 | ACTACCAGGTTCTGCTGAGGAGGGGTTTATAATCCATACTGGTTTCAACGCA | 360 |
| US-10-028-384-2 (1-826) x US-10-417-375-97 (1-5827) | | | | | | | |
| Qy | 70 | SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu | 89 | Qy | 429 | LeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIle | 448 |
| Db | 181 | ACACTTTGAAGCTTCTCATCTCTGTCATGGCTGCTGCTATATCTCTCCACTGCTG | 240 | Db | 1258 | CTCTATTACTGCTTTAGCAACCTGCTGAGCCCGGATTTTATCATCATGATGGTGTG | 1317 |
| Qy | 90 | PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheSerIleArg | 109 | Qy | 449 | SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal | 468 |
| Db | 241 | TTTGCTGCTGAGATTGAAGTGTATCCATGAGTTGATCCCGTACTTTAATATCGG | 300 | Db | 1318 | ACCAGCATGACTTTTTCAGCTGTAATGTCGCTAATGCTAGTGTGGCACCTGTTATG | 1377 |
| Qy | 110 | SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg | 129 | Qy | 469 | CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGluYAsp--- | 487 |
| Db | 301 | ACTACCAGGTTCTGCTGAGGAGGGGTTTATAATCCATACTGGTTTCAACGCA | 360 | Db | 1378 | TGCATTTCTCTGGCATGGAGTCTCCAGGTGCTGCCACATACATGAAGAACTGGAC | 1437 |
| Qy | 130 | AlaTrpTyrProLeuGluValGlyGlyThrValTyrProGlyLeuMetIleThr | 149 | Qy | 488 | MetLysArgGluLeuProValGluAspSerSerAspGluAspLysArgAsnGln | 507 |
| Db | 361 | GCTGTGACCTTTGGGAGCAATCATGGAGGAACTATTACCCAGGTTTATGATCACC | 420 | Db | 1438 | ATAAGTCGTCCA----- | 1449 |
| Qy | 150 | AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal | 169 | Qy | 508 | GlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThr | 527 |
| Db | 421 | TCTGCTGCAATCTACCATGCTACTCCATTTTCCACATCACCATCGACATCGGAATGTC | 480 | Db | 1450 | -----GACAAAGAGACGACAGACAGATTCACCTACCCCTATTAGAAT | 1497 |
| Qy | 170 | CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr | 189 | Qy | 528 | GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuLeu | 547 |
| Db | 481 | TGTTGTTCTCGCCCT | 540 | Db | 1498 | GAA-----GTGGCAAGTGGGATGATATGCTGCTCATGCTGCTCTCTCTCTCTCT | 1539 |
| Qy | 190 | ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro | 209 | Qy | 548 | MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerVal | 567 |
| Db | 541 | AAAGAGCTCAAGATGACGGGCTGGCTTCTTGTCTGCTGCTGCTGCTGCTGCTGCTG | 600 | Db | 1540 | ATCACCTACACCTTTCATTCACCTGGGTGACAGGTGAGGCTGCTCTCTCTCTCTCTCT | 1599 |
| Qy | 210 | GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla | 229 | Qy | 568 | ValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla | 587 |
| Db | 601 | GGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 660 | Db | 1600 | GTAATCTCCCGCTGCTGGGATGCGAGTAGGATCATATTTGATGACTTCCGAGAGCA | 1659 |
| Qy | 230 | LeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThr | 249 | Qy | 588 | TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr | 607 |
| Db | 661 | ATGCTACTCCTACTACTGATGGATCAGGAGGAAAGACATGGTTCCATCTGTGGGCA | 720 | Db | 1660 | TATATTGGCTCTCTCATATATCTCCAGAGGATGCGAAGGTCTGCTCTCTCTCTCTCTCT | 1719 |
| Qy | 250 | MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle | 269 | Qy | 608 | GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn | 627 |
| Db | 721 | GCTAAGTGTGCCCTTGTATTCTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 780 | Db | 1720 | GGCTATCAGATTACAGCTATGGCAACCCGAAACAATTTTATGGACATTAACATCGAAT | 1779 |

| Percent Similarity: | 74.44% | Conservative: | 103 |
|---|--------|---|------|
| Best Local Similarity: | 59.89% | Mismatches: | 145 |
| Query Match: | 50.92% | Indels: | 36 |
| DB: | 18 | Gaps: | 7 |
| US-10-028-384-2 (1-826) x US-10-417-375-99 (1-5404) | | | |
| QY | 70 | SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu | 89 |
| DB | 181 | ACACTTTTGAAGCTTCTCATCTGTCFCAATGGCTCTGTATTTATCTTCTCCACTCGGCTG | 240 |
| QY | 90 | PheAlaValIleArgPheCysIleIleHisGluPheAspProTrpPheAsnTyrArg | 109 |
| DB | 241 | TTTGCTGCTCGAGATTGAAAGTGTATCCATGAGTTGATCCGTACTTTAATTATCGG | 300 |
| QY | 110 | SerThrHisIleLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg | 129 |
| DB | 301 | ACTACCAAGTTCCTGGCTGAGGAGGGGTTTATAAATTCATAAATCGGTTTGTATGACCGA | 360 |
| QY | 130 | AlaTrpTyrProLeuGlyValGlyGlyThrValTyrProGlyLeuMetIleThr | 149 |
| DB | 361 | GGCTGTATCCCTTTGGAGCAATCATTGGAGAACATTTACCAGGTTTAATGATCCC | 420 |
| QY | 150 | AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal | 169 |
| DB | 421 | TCTGCTGCAATCTACCATGTACTCCATTTTCCACATCACCATCGCAATTCGGAATGTC | 480 |
| QY | 170 | CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr | 189 |
| DB | 481 | TGTGTGTTCTCGGCCCTCTCTCTCCCTTCCACCATCGTCACGTACCCACTTACC | 540 |
| QY | 190 | ArgGluLeuTrpAsnGlnGlyValGlyLeuLeuAlaAlaCysPheIleAlaIleValPro | 209 |
| DB | 541 | AAAGAGCTCAAGATGCAGGGCTGGGCTTCTGTCTGCATCATTTGCTGTAGTTCTCT | 600 |
| QY | 210 | GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla | 229 |
| DB | 601 | GGATATATCCCGATCTGTGGCTGGCTCCTATGATAATGAAGGATTTGCCATCTTTGC | 660 |
| QY | 230 | LeuGlnPheThrTyrTyrLeuTrpVallySerVallyThrGlySerValPheTrpThr | 249 |
| DB | 661 | ATGCTACTCACCTACTACATGTGAATGAGGACGAGTAAAGATCTGGTTCCATCTGTGGCA | 720 |
| QY | 250 | MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle | 269 |
| DB | 721 | GCTAAGTGTGCCCTTCTTATTCTACATGGTCTCGTCATGGGAGGTTATGTCTCTG | 780 |
| QY | 270 | IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg | 289 |
| DB | 781 | ATCAACTTAATCTCTCACGCTCTCGCTCGCTGCTGCTCACAGGCCGTTTCTCTCACCGG | 840 |
| QY | 290 | ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro | 309 |
| DB | 841 | ATCTATGTGGCTACTGTACTTGTACTGCTGGGCACTATATCTTCTATSCAGATCTCC | 900 |
| QY | 310 | PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla | 329 |
| DB | 901 | TTTGTGGGTTTCCAGGCTGTCTCTTTCATCAGACACATGGCAGCCGTTTGGGGTCTTTGGT | 960 |
| QY | 330 | LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe | 349 |
| DB | 961 | CTCTGCCAGATCATGCCCTTTGGGATTTACTTCCGCGCAGCAGTGAATCCACACAAATTT | 1020 |
| QY | 350 | GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal--- | 368 |
| DB | 1021 | GAAGTCTTTTCCGGAGCGTCACTCTCTG---GTAGGCTTTGTCTCTCTCACCGTGGGA | 1077 |
| QY | 369 | IleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrp | 388 |
| DB | 1078 | GCTCTCTCATGCTGACAGGAAATAATCTCTCTCGCGGGCGGTTTCTACTCGCTGCTG | 1137 |
| QY | 389 | AspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro | 408 |

210 GYTYRILSErArgSerValAlaGlySerPheAspAsnGluCylValAlaIlePheAla 229
Db GGATATATCTCCCATCTGTGGCTGCTATGATAAATGAAGGAGATTGCCATCTTTTGC 631
QY LeuGlnPheThrTYrLeuTrpValIysSerValIysThrGlySerValPheTrpThr 249
Db ATCTACTACCTACTACATCGGATCAAGGCAGTAAAGACTGGTTCTCATCTGTGGCA 691
QY MetCysCysLeuSerTYrPheTYrMetValSerAlaTrpGlyGlyValPheIle 269
Db GCTAAGTGTGCCCTTGCTATTCTCATCGTCTGTCATGGGAGGTATGTGTTCTCTG 751
QY IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTYrSerIysArg 289
Db ATCACTTAATTCCTCTCCACGCTCTGCTGCTCATGCTCACAGGCGGTTCTCACCGG 811
QY ValTYRILeAlaTYrSerThrPheTYrIleValGlyLeuIleLeuSerMetGlnIlePro 309
Db ATCTAATGTGGCTACTGTACTGTTTACTGCTGGGTACTATATCTTCTTAGCGAGATCTC 871
QY PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
Db TTTGTGGTTCACAGCGCTGCTTTCATCATGAGACATGCGAGGTTTGGGCTCTTGGT 931
QY LeuLeuGlnAlaTYrAlaPheLeuGlnTYrLeuArgAspArgLeuThrIysGlnIlePhe 349
Db CTCTGCCAGATCCAGCTTTTGGATTACTCTGGCAGCAGTTGAATCCACACAAATTT 991
QY GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal--- 368
Db GAAGTCTCTTTCGAGAGCGTCATCTCTCTG---GTAGGCTTTGTCTTCTCACGCTGGGA 1048
QY IleTYrLeuThrTYrThrGlyTYrIleAlaProTrpSerGlyArgPheTYrSerLeuTrp 388
Db GCTCTCTCTGCTGCTACAGGAAAATAATCTCCCTGGACGGGGCTTTCTACTCATCTGCTG 1108
QY AspThrGlyTYrAlaLYsIleHisIleProIleIleAlaSerValSerGluHisGlnPro 408
Db GATCCCTCTTATGCTAAGAACCAATCCCATCATTTGCTCTGTGCTGTGAGCATCAGCC 1168
QY ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
Db ACAACCTGGTCTCATACTATTTTGACTCGAGCTCTGCTTTCATGTTTCCAGTTGGC 1228
QY LeuTrpPheCysIleLYsAsnIleAspAspGluArgValPheValAlaLeuTYrAlaIle 448
Db CTCTATTACTGCTTTAGCAACCTCTGTGATGCCCGGATTTTATCATCATGTATGGTGTG 1288
QY SerAlaValTYrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal 468
Db ACCAGCATGTACTTTTCAGCTGTAAATGGTGGCTTAATGCTAGTGTGGCACCCTGTATG 1348
QY CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTYrIleuGlyAsp---Asp 487
Db AGCATTTCTCTGGCATTTGGAGTCTCCACAGTCTCTCCACATACATCAAGAATCTGCAC 1408
QY MetLYsArgGluAsnProProValGluAspSerSerAspGluAspAspIysArgAsnGln 507
Db ATAACTCGCCCA----- 1420
QY GlyAsnLeuTYrAspLYsAlaGlyLYsValArgLYsHisAlaThrGluGlnGlyThr 527
Db -----GACAAAGACGACGAAGCAACAGAGATTCCACCTTACCTTATTAAAGATT 1468
QY GluGluGlyLeuGlyProAsnIleLYsSerIleValThrMetLeuMetLeuMetLeu 547
Db -----GTGGCAAGTGGGATGATACTGTCATGCTTCTTCTTCTC 1510
QY MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTYrSerSerProSerVal 567
Db ATCACCATCACTTTTCAATCAACTCGGTGACAGCTGAGCCCTACTCTTCTCCGCTCCAT 1570
QY ValLeuAlaSerTYrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla 587

| | | | |
|----|--------|---|------|
| Db | 1571 | GTACTATCTGCCCGTGGGATGGCAGTAGGATCAATATTGATCACTTCGAGAGCA | 1630 |
| Qy | 588 | TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr | 607 |
| Db | 1631 | TATTATTGGCTTCGTATATACTCTCAGAGAGATGCGAAGTCATGCTCTGGTGGGATTAT | 1690 |
| Qy | 608 | GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn | 627 |
| Db | 1691 | GGCTATCAGATTACAGCTATGGCCACCCGACCAATTTTAGTCGACAATAACACATGGAA | 1750 |
| Qy | 628 | AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrIlys | 647 |
| Db | 1751 | AATACCCCATTTCTCCAGTAGGCGCAGGCAATGCGCTCCACAGAGCAAAAAGCCTATTGAG | 1810 |
| Qy | 648 | IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr | 667 |
| Db | 1811 | ATCATGGGGAGCTCGATGTCAGCTATGTCGTCTCATTTTGGAGGCTCACTGGGTAT | 1870 |
| Qy | 668 | SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAla-----Glu | 683 |
| Db | 1871 | TCTCTGATGATATACCAAGTTCTTTGGATGGTCCGATTTGGAGGGAGCACAGATACA | 1930 |
| Qy | 684 | GlyGluHisProLysAspIleArgLysSerAspTyrPheThrProGlnGlyGluPheArg | 703 |
| Db | 1931 | GGCAACAT-----ATCAAGGAGAATGACTATTATACTCCAACCTGGGAGGTCCGT | 1981 |
| Qy | 704 | ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrIlysMetSerTyrTyr | 723 |
| Db | 1982 | GTGGACCTTGAAGGTTCTCCAGTCTGCTCAACTGCTCATGTACAGAGTGTTTACTAT | 2041 |
| Qy | 724 | ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn | 743 |
| Db | 2042 | CGTTTGGACAGGTTTACACAGACCCACGCGCTCCAGGCTTTGACCGTCTCCGAAT | 2101 |
| Qy | 744 | AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaAlaPheThrSerGlu | 763 |
| Db | 2102 | GCTGAGATTGGGAATAAAGACTTTGACGTTGATGTCCTGTGAGAAAGGCTTATACCAAGAA | 2161 |
| Qy | 764 | HisTrpLeuValArgIleTyrLysValIysAlaProAspAsnArg-GluThrLeuAspHi | 783 |
| Db | 2162 | CATTGGCTGGTCAGGATATACAAGGTAAGGACCTGGATAATCGAGGCTTGTCAAGGACA | 2221 |
| Qy | 783 | SysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrLys | 803 |
| Db | 2222 | TAAATGTACGTCACGCTCTGATATCTTCGCACTGAGCACATCACATTTAGGACCTTGA | 2281 |
| Qy | 803 s | 803 | |
| Db | 2282 g | 2282 | |

RESULT 11

```

US-10-417-375-99
; Sequence 99, Application US/10417375
; Publication NO. US20040219528A1
; GENERAL INFORMATION:
;
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 99
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
;

```

| | | |
|-------------------|-----------|--------------|
| Alignment Scores: | | |
| pred. No.: | 1.59e-221 | Length: 5404 |
| Score: | 2241.00 | Matches: 424 |

Db 1229 CTCATTACTGCTTTAGCAACCTGCTGATGCCCGGATTTTATCATCATGTATGGTGTG 1288
 Qy 449 SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal 468
 Db 1289 ACCAGATGATCTTTTACGCTGTATGGTGGTCTAATGCTAGTGTGGCACCCTGTTATG 1348
 Qy 469 CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAsp---Asp 487
 Db 1349 AGCAATCTCTCTGGCATTGGAGTCTCCAGAGTGGTGTCCACATACATGAAGAATCTGGAC 1408
 Qy 488 MetLysArgGluAsnProValGluAspSerSerAspGluAspAspLysArgAsnGln 507
 Db 1409 ATAAGTCGCCCA----- 1420
 Qy 508 GlyAsnLeuTyrAspLysAlaGlyValArgLysHisAlaThrGluGlnGluLysThr 527
 Db 1421 -----GACAAAGAGCAAGCAAGCAACAGGATTCACCTACCTCTATTAAAGATT 1468
 Qy 528 GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeu 547
 Db 1469 GAA-----GTGGCAAGTGGGATGATCTGTCATGGCTTCTTCTC 1510
 Qy 548 MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerVal 567
 Db 1511 ATCACTACACCTTTTCATTCACCTGGGTGACAGTGAGGCTCTCTCTCCGTCCTAT 1570
 Qy 568 ValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla 587
 Db 1571 GTACTATCTGCCGTGGTGGGATGCGAGTAGATCATATTTGATGACTTCGGAAGCA 1630
 Qy 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpAspTyr 607
 Db 1631 TATATGCTTCTGCTCATATCTCCAGAGGATGCGAGGTCATGTCCTGGTGGGATTAT 1690
 Qy 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
 Db 1691 GGCTATCAGATTACAGCTATGCAAAACCGAATAATTTTATGGACAAATAACATGGGAAT 1750
 Qy 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
 Db 1751 AATACCATATTTCTCGAGTGGGAGGCAATGGCGTCCAGAGGAAAAAGCCTATGAG 1810
 Qy 648 IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyr 667
 Db 1811 ATCATGAGGAGCTCGATGTCAGTATGCTGCTGCTCATTTTGGAGGCTCCTACCTGGTAT 1870
 Qy 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAla-----Glu 683
 Db 1871 TCCTCTGATGATATCAACAAGTTTCTTTGGATGGTCCGGATTGGAGGAGCACAGATACA 1930
 Qy 684 GlyGluHisProLysAspIleAsgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 Db 1931 GGCACAACT-----ATCAGGAGATGACTATATATCTCAACTCGGAGGATTCCT 1981
 Qy 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 Db 1982 GTGACCGTGAAGGTTCTCCAGTCTCTCAACTGCTCATGTACAGATGTGTTACTAT 2041
 Qy 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
 Db 2042 CGCTTTGGACAGGTTTACAGAGCAAGCCAGCTCTCCAGGCTTTGACCGTGTCCGAAT 2101
 Qy 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGlu 763
 Db 2102 GCTGAGATTGGGAATAAGACTTTGAGCTTGTATGCTCTGGAGGAGGCTATACACAGAA 2161
 Qy 764 HistTrpLeuValArgIleTyrLysValIleAlaProAspAsnArg-GluThrLeuAspH 783
 Db 2162 CATTTGGCTGGTACAGATATACAGATTAAGAGCTGTGATATCGAGGCTGTCAAGGACA 2221
 Qy 783 sLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLy 803
 Db 2222 TAAATGTCCAGCTCTGATATCTTCGCACTGAGCACATCACATTTAGGAGCTGTGA 2281

Qy 803 s 803
 Db 2282 G 2282
 RESULT 10
 US-10-342-887-742
 ; Sequence 742, Application US/10342887
 ; Publication No. US20040058340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter S.
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Christopher J.
 ; APPLICANT: Van t Veer, Laura Johanna
 ; APPLICANT: Van de Vijver, Marc J.
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-188-999
 ; CURRENT APPLICATION NUMBER: US/10/342,887
 ; CURRENT FILING DATE: 2003-01-15
 ; PRIOR APPLICATION NUMBER: 60/298,918
 ; PRIOR FILING DATE: 2001-06-18
 ; PRIOR APPLICATION NUMBER: 60/380,710
 ; PRIOR FILING DATE: 2002-05-14
 ; PRIOR APPLICATION NUMBER: 10/172,118
 ; PRIOR FILING DATE: 2002-06-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-342-887-742
 Alignment Scores:
 Pred. No.: 5,16e-223 Length: 2472
 Score: 2250.00 Matches: 432
 Percent Similarity: 73.01% Conservative: 109
 Best Local Similarity: 58.30% Mismatches: 163
 Query Match: 51.12% Indels: 37
 DB: 16 Gaps: 7

US-10-028-384-2 (1-826) x US-10-342-887-742 (1-2472)
 Qy 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu 89
 Db 152 ACATTTTGAAGCTTCTCATTTCTGTCATAGGCTGCTATTATTCCTTCTCCACTCGTCTG 211
 Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
 Db 212 TTTGCTGCTCGAGATTGAAGTGTATCATGAGTTGATCGTACTTTATTTATTCGG 271
 Qy 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
 Db 272 ACTACCGAGTTCTCGCTGAGGAGGGGTTTATAAATCCATAACTGCTGTTGATGACCGA 331
 Qy 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyValTyrProGlyLeuMetIleThr 149
 Db 332 GCCTGGTACCTTTGGGACGAATCATTTGGAGGAACAATTTACCAGGTTTATGATCACC 391
 Qy 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 Db 392 TCTGCTGCAATCTACCATGTACTCCATTTTTCACATCACCACGACATTCGGAATGTC 451
 Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 452 TGTGTGTCTCTGGCCCT 511
 Qy 190 ArgGluLeuTrpAsnGlnGlyValGlyLeuAlaAlaCysPheIleAlaIleValPro 209
 Db 512 AAGAGCTCAAGGATGAGGGGCTGGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 571

QY 668 SerGlyAspAlaLeuSerPheLeuThrMetValArgIleAla-----Glu 683
 DB 1871 TCCTCTGATGATATCAACAAGTTCTTTGGATGGTCCGGATTGGAGGACACAGATACA 1930
 QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 DB 1931 GGCACCAAT-----ATCAAGAGAGATGACTATTATATCTCAACTGGGGAGTTCCGT 1981
 QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 DB 1982 GTGACCGTGAAGGTTCTCCAGTCTCTCACTGCTCACTGCTCACTGACAGATGTTACTAT 2041
 QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
 DB 2042 CGCTTTGGACAGGTTTACACAGAAGCCAGCGCTCTCCAGGCTTTCACCGTGTCGAAAT 2101
 QY 744 AlaGluIleGlyLeuLysAspIleLysPheLysHisLeuIleGluAlaPheThrSerGlu 763
 DB 2102 GCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCTTGGAGGAAGCTATACACAGAA 2161
 QY 764 HistPheValArgIleTyrLysValLysAlaProAspAsnArg-GluThrLeuAspHi 783
 DB 2162 CATTTGGCTGTGAGGATATACAAAGTAAAGACCTGGATAATCGAGGCTTGTCAAGGACA 2221
 QY 783 sLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysThrThrly 803
 DB 2222 TAAATGTCAGCTCAGCTCTGATATCTTCCACTGAGCACATCACATTTAGGACGTTGAA 2281
 QY 803 s 803
 DB 2282 G 2282

RESULT 9

US-10-172-118-742
 ; Sequence 742, Application US/10172118
 ; Publication No. US20030224374A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yundong
 ; APPLICANT: Linsley, Peter
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Chris
 ; APPLICANT: Van 't Veer, Laura
 ; APPLICANT: Van de Vijver, Marc
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-135-999
 ; CURRENT APPLICATION NUMBER: US/10/172,118
 ; PRIOR FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: 60/380,770
 ; PRIOR FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_002219
 ; DATABASE ENTRY DATE: 2001-06-18
 US-10-172-118-742

Alignment Scores:

| Pred. No.: | Score: | Length: | Matches: |
|------------------------|---------|---------------|----------|
| 5.16e-223 | 2250.00 | 2472 | 432 |
| Percent Similarity: | 73.01% | Conservative: | 109 |
| Best Local Similarity: | 58.30% | Mismatches: | 163 |
| Query Match: | 51.12% | Indels: | 37 |
| DB: | 15 | Gaps: | 7 |

US-10-028-384-2 (1-826) x US-10-172-118-742 (1-2472)

QY 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu 89

DB 152 ACACATTTGAAGCTTCTCATTTCTGCAATGGCTGCTGTATTATATCTTCTCCACCTCGTGTG 211
 QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
 DB 212 TTTTGTCTCTCAGAGATTTGAAAGTGTATCATCAGATTGTATCGTACTTTAATTAATCGG 271
 QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
 DB 272 ACTACACAGTTCTTGGCTGAGGAGGGTTTATTAATTTCCATAACTGGTTGATGACCGA 331
 QY 130 AlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThr 149
 DB 332 GCTGTGTACCTTTGGGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACC 391
 QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 DB 392 TCTGTGCAATCTACCATGTACTCCATTTTCCACATCCACCATTCGAAATGTC 451
 QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 DB 452 TGTGTGTCTTGGCCCT 511
 QY 190 ArgGlnLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
 DB 512 AAAGAGCTCAAGGATGAGGGGCTGGGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 571
 QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 DB 572 GGATATATCTCCCATCTGTGGCTGGCTCTATGATTAATGAAGGATTCGCACTTTTGC 631
 QY 230 LeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
 DB 632 ATGCTACTACCTACTATCATGTGATCAAGCAGTAAGACTGGTTCCATCTCTGTGGCA 691
 QY 250 MetCysCysLysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
 DB 692 GCTAAGTGTGCCCTTGGTTTCTATGCTCTGTCATGGGAGGTTAATGTTGTTCTGT 751
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
 DB 752 ATCAACTTAATCTCTCCAGCTCTCTGCTGATGTGTACAGCGCGTTTCTCTCACCAG 811
 QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 DB 812 ATCTATGTGGCTTACTGTACTGTTTACTGCTGGGACTATATCTTCTTAGCGAGATCTCC 871
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
 DB 872 TTTGTGGGTTTCCAGCGCTGCTTTCATCAGACACATGCGAGGTTTGGGGTCTTTGGT 931
 QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
 DB 932 CTCTGCCAGATCCATGCCTTTGTGGATTACTCTCGCAGCAAGTGAATCCACACAATTT 991
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyValPheLeuSerVal--- 368
 DB 992 GAAGTCTTTTCCGGAGCGTCATCTCTG---GTAGGCTTTTCTCTCTCTCTCTCTCTCT 1048
 QY 369 IleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrp 388
 DB 1049 GCTCTCTCATGTGTGACAGGAAAAATATCTCCCTGGACGGGGCTTCTACTCACTGTCTG 1108
 QY 389 AspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnPro 408
 DB 1109 GATCCCTCTTATGCTAAGAACAAATCCCATCATCTGCTCTGTGCTGAGGATCAGCCC 1168
 QY 409 ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
 DB 1169 ACAACCTGGTCTCATATTTTGGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1228
 QY 429 LeuTrpPheCysIleLysAsnIleAsnArgGluArgValPheValAlaLeuTyrAlaIle 448

APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028,384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent in version 3.1
SEQ ID NO 11

LENGTH: 2472
TYPE: DNA
ORGANISM: Homo sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: NM_002219
DATABASE ENTRY DATE: 2000-10-31
RELEVANT RESIDUES: (1)..(2472)
US-10-028-384-11

Alignment Scores:
Pred. No.: 5,166-223 Length: 2472
Score: 2250.00 Matches: 432
Percent Similarity: 73.04% Conservatives: 109
Best Local Similarity: 58.30% Mismatches: 163
Query Match: 51.12% Indels: 37
DB: 15 Gaps: 7

US-10-028-384-2 (1-826) x US-10-028-384-11 (1-2472)

| | | | |
|----|-----|--|-----|
| QY | 70 | SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu | 89 |
| DB | 152 | ACACTTTGAAGCTTCTCATCTCTGTCATGGCTGCTGTTATTTATCTCTCCACTCGTCTG | 211 |
| QY | 90 | PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg | 109 |
| DB | 212 | TTTGCTGCTCGAGATTGAAGTGTATCCATGAGTTGATCCCGTACTTTAATTATCGG | 271 |
| QY | 110 | SerThrHisHisLeuAlaSerHisGlyPheTrpGluPheLeuAsnTrpPheAspGluArg | 129 |
| DB | 272 | ACTACCAAGTTCCTGGCTGAGGAGGGTTTATAAATTCATAAATCTGGTTTCATCACCGA | 331 |
| QY | 130 | AlaTrpTrpProLeuGlyArgIleValGlyThrValTrpProGlyLeuMetIleThr | 149 |
| DB | 332 | GCCTGTGTACCTTTGGGAGCAATCATTTGGAGGAAATTTACCCAGGTTTATATGATCAC | 391 |
| QY | 150 | AlaGlyLeuIleHisTrpIleLeuAsnTrpLeuAlaAlaCysPheIleAlaIleVal | 169 |
| DB | 392 | TCGTGCAATCATCATGATCTCCATTTTCCATCATCACATCGACATTCGGATGTC | 451 |
| QY | 170 | CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuThr | 189 |
| DB | 452 | TGTGTGTTCTGGCCCT | 511 |
| QY | 190 | ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro | 209 |
| DB | 512 | AAAGAGCTCAAGATGAGGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 571 |
| QY | 210 | GlyTrpIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla | 229 |
| DB | 572 | GGATATATCTCCGATCTGTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 631 |
| QY | 230 | LeuGlnPheThrTrpTrpValIleSerValIleThrGlySerValPheThrThr | 249 |
| DB | 632 | ATGTACTACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT | 691 |
| QY | 250 | MetCysCysLeuSerTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPheIle | 269 |
| DB | 692 | GCTAAGTGTGCCCTTGTCTATTTCTATCTATCTATCTATCTATCTATCTATCTATCT | 751 |
| QY | 270 | IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTrpSerIleArg | 289 |
| DB | 752 | ATCAACTAATCTCTCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 811 |
| QY | 290 | ValTrpIleAlaTrpSerThrPheTrpIleValGlyLeuIleLeuSerMetGlnIlePro | 309 |

| | | | |
|----|------|--|------|
| DB | 812 | ATCTATGTGGCTCTACTGTTACTGCTGCTGGGTACTATATCTTCTTAGGCAGATCTCC | 871 |
| QY | 310 | PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla | 329 |
| DB | 872 | TTTGTGGGTTCCAGCT | 931 |
| QY | 330 | LeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeuArgAspArgLeuThrLysGlnGluPhe | 349 |
| DB | 932 | CTCTGCAGATCATCCCTTTGTGGATTTACCTTGGCAGCAAGTTGAATCCACAATTT | 991 |
| QY | 350 | GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal | 368 |
| DB | 992 | GAGTCTTTTTCGGAGCGCTCATCTCTCTG---GTAGGCTTTTCTCTCTACCGCTGGA | 1048 |
| QY | 369 | IleTrpLeuThrTrpThrGlyTrpIleAlaProTrpSerGlyArgPheTrpSerLeuTrp | 388 |
| DB | 1049 | GCTCTCTCTCATGCTGACAGGAAATATCTCTCTGAGCGGGCGTTTCTACTCATGCTG | 1108 |
| QY | 389 | AspThrGlyTrpAlaLysIleHisIleProIleAlaSerValSerGluHisGlnPro | 408 |
| DB | 1109 | GATCCCTCTTATGCTAAGAACACATCCCATCATTTCTGTTCTGTGCTGAGCATCAGCC | 1168 |
| QY | 409 | ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly | 428 |
| DB | 1169 | ACAACTGTGCTCTACTATTATTTGACCTGACGCTCTCTCTCTCTCTCTCTCTCTCT | 1228 |
| QY | 429 | LeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTrpAlaIle | 448 |
| DB | 1229 | CTCTATTTACTGCTTTAGCAACATCCCATCATTTCTGTTCTGTGCTGAGCATCAGCC | 1288 |
| QY | 449 | SerAlaValTrpPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal | 468 |
| DB | 1289 | ACCAGCATGTACTTTTTCAGCTGTAATGCTGCTCTAATGCTAGTGTGGCCTCTGTATG | 1348 |
| QY | 469 | CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTrpLeuGlyAsp | 487 |
| DB | 1349 | AGCATTTCTCTGGCATGAGTCTCCAGGTCTCTCCATCATCATGATGAGATCTGGAC | 1408 |
| QY | 488 | MetLysArgGluAsnProValGluAspSerSerAspGluAspLysArgAsnGln | 507 |
| DB | 1409 | ATAAGTCGCCCA----- | 1420 |
| QY | 508 | GlyAsnLeuTrpAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThr | 527 |
| DB | 1421 | -----GACAAAGACAGCAAGCAAGCAAGATTTCCACCTACCTATTAGATT | 1468 |
| QY | 528 | GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeu | 547 |
| DB | 1469 | GAA-----GTGGCAAGTGGGATGATGATGCTGCTGCTGCTCTCTCTCTCTCT | 1510 |
| QY | 548 | MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTrpSerSerProSerVal | 567 |
| DB | 1511 | ATCACCTACACCTTTTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1570 |
| QY | 568 | ValLeuAlaSerTrpAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla | 587 |
| DB | 1571 | GTACTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1630 |
| QY | 588 | TrpPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTrp | 607 |
| DB | 1631 | TATTATTGGCTTCTGCTATATATCTTCCAGAGATGCGAAGGCTGCTGCTGCTGCTGCT | 1690 |
| QY | 608 | GlyTrpGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn | 627 |
| DB | 1691 | GGCTATCATGATATGAGTATGGCAACCGAACAATTTTATGACATATACATGGAAT | 1750 |
| QY | 628 | AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGlnThrAlaAlaTrpLys | 647 |
| DB | 1751 | AATACCATATTTCTCGAGTAGGCGCAATGCGTCCACAGAGGAAAAAGCGCTAGAG | 1810 |
| QY | 648 | IleMetArgThrLeuAspValAspTrpValLeuValIlePheGlyValIleGlyTrp | 667 |
| DB | 1811 | ATCATGAGGAGCTCGATGTCAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1870 |

332 GCCTGGTACCGCTTTGGGACGAAATCATTTGGGAGCAAAATTTACCCAGGTTTAAATCATCACCC 391
 150 AlaGlyLeuIleHisThrPheLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 392 TCTGCTGCAATCTACCATGTAATCTTCTCCACATCACCATCCACATTCGGAATGTC 451
 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 452 TGCTGTGCTCCGGCCCT 511
 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 512 AAAGAGCTCAAGGATGAGGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 571
 210 GlyThrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 572 GGAATATCTCCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 631
 230 LeuGlnPheThrTyrrTrpLeuTrpValIleSerValIleThrGlySerValPheThrThr 249
 632 ATGCTACTCACCTACTACATGTGATCAAGGAGTAAAGAGCTGTTCCATCTCTGTTGGCA 691
 250 MetCysCysLeuSerTyrrPheTyrrMetValSerAlaTrpGlyTyrrValPheIle 269
 692 GCTAAGTGTGCCCTTGCTTATTTCTACATGCTCTGCTCATGGGAGGTTATGTTCTCTG 751
 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrrSerIleArg 289
 752 ATCAACTTAAATCTCTCTCCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811
 290 ValTyrrIleAlaTyrrSerPheTyrrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 812 ATCTATGGGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 871
 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 872 TTGTGGGTTTCCAGCTGCTCTTTCATCAGAGCACATGCGAGGTTTGGGCTTTTGGT 931
 330 LeuLeuGlnAlaTyrrAlaPheLeuGlnTyrrLeuArgAspArgLeuThrIleGluPhe 349
 932 CTCTGCCAGATCCATGCTTGTGGATTACCTGCGCAGCAAGTTGAATCCACAACTT 991
 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaGlyAlaValPheLeuSerVal--- 368
 992 GAAGTCTCTTCCGAGCGTCTCTCTCTG---GTAGGCTTTCTCTCTCTCTCTCTCTCTCTCT 1048
 369 IleTyrrLeuThrTyrrThrGlyTyrrIleAlaProTrpSerGlyArgPheTyrrSerLeuTrp 388
 1049 GCTCTCTCTCATGCTGACAGGAAATAATCTCTCTGAGCGGGGCTTCTACTCATCTGCTG 1108
 389 AspThrGlyTyrrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro 408
 1109 GATCCCTCTTATGTAAGAACACATCCCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1168
 409 ThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGly 428
 1169 ACAACCTGCTCTCATCTATTTTGGACCTGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1228
 429 LeuTrpPheCysIleLeuAsnIleAsnAspGluArgValPheValAlaLeuTyrrAlaIle 448
 1229 CTCATTACTGCTTAGCAACCTCTCTGATGCTCCCGGATTTTATCATCATGATGATGCTG 1288
 449 SerAlaValTyrrPheAlaGlyValMetValArgLeuLeuLeuThrLeuThrProValVal 468
 1289 ACCAGCATGTACTTTTCCAGCTGTAATGGTGGCTCTAATGCTAGTGTGGACCTGTATG 1348
 469 CysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrrLeuGlyAsp---Asp 487
 1349 AGCATTTCTCTGGAATTTGGAGTCTCCAGGTGCTGCTCCACATACATGAAGATCTGGAC 1408
 488 MetIleLysGluAsnProProValGluAspSerSerAspGluAspIleAspValArgAsnGln 507

1409 ATAAGTCGCCCA----- 1420
 508 GlyAsnLeuTyrrAspIleValArgLysValArgLysHisAlaThrGluGlnGluIleThr 527
 1421 -----GACAAGAAGAGCAAGCAAGCAAGGATTTCACCTACCTATTAAAGATT 1468
 528 GluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuLeu 547
 1469 GAA-----GTGGCAAGTGGGATGATCTATGCTATGCTGCTCTCTCTCTCTCTCTCTCT 1510
 548 MetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrrSerSerProSerVal 567
 1511 ATCACTTACACCTTTTCACTCAACCTGGGTGACCAAGTGGAGCTCTCTCTCTCTCTCTCTCT 1570
 568 ValLeuAlaSerTyrrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla 587
 1571 GTACTATCTCCCGTGGTGGGATGCAATGATGATCATTTTGTGAGTCTTCGGAAGCA 1630
 588 TyrrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrr 607
 1631 TATTATTGGCTTCTCTATAATCTCTCAGAGGATGCGAGGTCATGCTCTCTCTCTCTCTCTCTCT 1690
 608 GlyTyrrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
 1691 GGTATCAGATTACAGCTATGCGCAACCAACCAATTTTAGTGGACATTAACATGGAAT 1750
 628 AsnSerHisIleAlaLeuValGlyIleAlaMetSerSerAsnGluThrAlaAlaTyrrLys 647
 1751 AATACCCATATTTCTCGAGTAGGGCAGGCAATGGGCTCCACAGAGGAAAGACCTATGAG 1810
 648 IleMetArgThrLeuAspValAspTyrrValLeuValIlePheGlyValIleGlyTyrr 667
 1811 ATCATGGAGGAGCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1870
 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAla-----Glu 683
 1871 TCCTCTCATGATATCAACAAGTTTCTTGGATGGTCCGGATTGGAGGGAGCACAGATACA 1930
 684 GlyGluHisProLysAspIleArgGluSerAspTyrrPheThrProGlnGlyGluPheArg 703
 1931 GGCACAACT-----ATCAAGGAGATGACTATTTACTTCCAACTGGGAGTTCCTG 1981
 704 ValAspIleAlaGlySerProThrLeuLeuAsnCysLeuMetTyrrLysMetSerTyrrTyrr 723
 1982 GTGACCGTCAAGGTTCTCCAGTGTCTCAACTGCTCATGTACAGATGTGTACTAT 2041
 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsn 743
 2042 CGCTTTGACAGGTTTACAGANGCAAGCGCTCTCCAGGCTTTGACCGGTTCGGAAT 2101
 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGlu 763
 2102 GCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCTCGAGGAGGCTATACCAAGAA 2161
 764 HisTrpLeuValArgIleTyrrLysValLysAlaProAspAsnArg-GluThrLeuAspHis 783
 2162 CATTTGGTGTCTAGGATATACAGGTAAGGACCTGGATTAATCGAGGCTTGTCAAGGACA 2221
 783 SLysProArgValThrAsnIlePheProLysGlnIleTyrrLeuSerLysLysThrThrly 803
 2222 TAAATGTCACTCCAGCTCTGATATCTTCGCACTGAGCACATCACATTTAGGACGTTGAA 2281
 803 a 803
 2282 G 2282
 RESULT 8
 US-10-028-384-11
 ; Sequence 11, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude

QY 439 GluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetVal 458
 Db 1264 GTCGGATTTTATCATCATGATATGGTGTGACACGATGACTTTTCAGCTGTAATGGTG 1323
 QY 459 ArgLeuMetLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAen 478
 Db 1324 CQTCTAATGCTGGTATTGGCACCTGTTATGTGCATCTTTCTGGCATGGTGTTCCTCCAG 1383
 QY 479 ValPheGluHisTyrLeuGlyAsp---AspMetIlySargGluAenProProValGluAap 497
 Db 1384 GTGCTGTCCACATATATGAAATATCGACATAAGTCGCCCA-----1425
 QY 498 SerSerAspGluAspAspIlyArgAenGlnGlyAenLeuTyrAspIlySalaGlyIlySVal 517
 Db 1426 -----GACAAAGAACGACAAAG 1443
 QY 518 ArgIlyHisAlaThrGluGlnGluIlyThrGluGluGlyLeuGlyProAenIlySarg 537
 Db 1444 CACAGGATTTCTACTTCCCTATTAAAGATGAG-----GTGGCGAGT 1485
 QY 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 Db 1486 GGGATGATACCTGGCTGTTTCTCATCACCTACACGTTTCTTCGACTTGGGTG 1545
 QY 558 ThrSerAenAlaTyrSerSerProSerValValLeuAlaSerTyrAenHisAspGlyThr 577
 Db 1546 ACCAGTGAGGCTATTTCTTCTCTCATGTACTGCTCTGCTGGTGGGATGGCAT 1605
 QY 578 ArgAenIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAenThrAspGlu 597
 Db 1606 AGGATCATTTTGTATGACTTCCGAGAGCGTATTATTGGTCTCGTCCATCACATATCCACAG 1665
 QY 598 HisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAenArg 617
 Db 1666 GATGCAAGATCATGTATGGTGGATATGGCTACCAATTAATCTGCAATGGCAATCG 1725
 QY 618 ThrThrLeuValAspAenAsnThrTrpAsnAenSerHisIleAlaLeuValGlyIlySala 637
 Db 1726 ACAATTTTAGTGGACAAATAACATGATGAATAATACCCATATTCTCGAGTAGGCGAGCA 1785
 QY 638 MetSerSerAenGluThrAlaAlaTyrIlyIleMetArgThrLeuAspValAspTyrVal 657
 Db 1786 ATGGATCCACAGAGAAAGCCCTATGAATATATGAGGAGCTTGATGTACGTATGTG 1845
 QY 658 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsnIlySalaLeuTrp 677
 Db 1846 CTGTGCTATTTGGAGGCTTACTGGGTATTTCTCGGATGATATCAACAAGTTTCTTTGG 1905
 QY 678 MetValArgIle---AlaGluGlyGluHisProIlySargIleArgGluSerAspTyrPhe 696
 Db 1906 ATGGTCCGATTTGGAGNAGCACAGACAGACAGACACATTAAGGAGATGACTACTAT 1965
 QY 697 ThrProGlnGlyGluPheArgValAspIlyAlaGlySerProThrLeuLeuAenCysLeu 716
 Db 1966 ACTCTACTTGGGAATTCGGTGTGATCGTGAGGGTTCCTCGGTGCTGCTCAACTGCCCT 2025
 QY 717 MetTyrIlyMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
 Db 2026 ATGTACAAATGTGTTACTACCGTTTGGGAGGTCTACACAGACCAAGCGTCCACCA 2085
 QY 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAenIlySargIlySalaLeu 756
 Db 2086 GGTCTTGACCGTGTTCGAAATGCTGAGATTGGTAAATAAGACTTTGAGTTGATGCTG 2145
 QY 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrIlySalaProAsp 776
 Db 2146 GAGNAGCGTATACACAGAACTGGCTAGTCAGGATATACAGGTAAGGACCTGGAT 2205
 QY 777 AsnArg 778
 Db 2206 AATCGA 2211

RESULT 6

US-10-417-375-95
 ; Sequence 95, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 95
 ; LENGTH: 3046
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-417-375-95
 Alignment Scores:
 Pred. No.: 4,09e-224 Length: 3046
 Score: 2262.00 Matches: 429
 Percent Similarity: 74.41% Conservative: 103
 Best Local Similarity: 60.00% Mismatches: 147
 Query Match: 51.40% Indels: 35
 DB: 18 Gaps: 7
 US-10-028-384-2 (1-826) x US-10-417-375-95 (1-3046)
 QY 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeu 89
 Db 181 ACATTTTGAAGCTTCTCATTTCTGTCATATGGCTGCTGTTATTCCTTCTCCACTGCTG 240
 QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
 Db 241 TTGCTGCTCTGAGATTGAAAGTGTATCCAGAGTTTGATCCGACTTTAATATATCGG 300
 QY 110 SerThrHisIleAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
 Db 301 ACTACCAAGTCTCTGGCTGAGGAGGGTATTATATAATTCATACTGGTTTCATGACCGA 360
 QY 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
 Db 361 GCCTGGTACCTTTGGAGCAATCATGGAGAACAAATTTACCCAGGTTTAATGATCACC 420
 QY 150 AlaGlyLeuIleHisTrpIleLeuAenThrLeuAsnIleThrValHisIleArgAspVal 169
 Db 421 TCTGCTCAATCTACCATGACTCTCCATTTTTCACATCATCACCATCGACATTGGAAATGTC 480
 QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 481 TGTGTGTCTCTGGCCCTCTCTTCTCTCTCTCCATCCACCATCGTCACGATCCACCTTACC 540
 QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 Db 541 AAAGAGCTCAAGATGACAGGGGCTGGGCTTCTGCTGCTGCCATGATTGCTGTAGTTCT 600
 QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAenGluGlyIleAlaIlePheAla 229
 Db 601 GGATATATCTCCGATCTGTGGCTGCTCTCTATGATGAATGAAGGATGATGATGCTTTTGC 660
 QY 230 LeuGlnPheThrTyrLeuTrpValIlySerValIlySerValPheTrpThr 249
 Db 661 ATGCTACTCACTTACTTACTGATGATCAAGGAGTAAAGACTGGTTCCATCTGTGTGGCA 720
 QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
 Db 721 GCTAAGGTGCTCCCTTCTTCTATCTGATGGTCTGCTATGGGAGGTTATGTGTCTG 780
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerIlySarg 289
 Db 781 ATCAACTTAATCTCTCCACGCTCTGCTGCTGCTGCTCACAGGCGGTTCTCTCACCGG 840
 QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309

Qy 658 LeuValIlePheGlyGlyValIleGlyTyrSerClyAspAspIleAsnIleValPheLeuTyr 677
 Db 1846 CTTGTGCAATTTTGGAGGCTTACTGGGTATCTTCGGATGATATCAACAGTTCTTTGG 1905
 Qy 678 MetValArgIle---AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
 Db 1906 ATGGTCCGATGGAGGAAGCACAGACAGAGGAAGACACATTAAAGGAGATGACTACTAT 1965
 Qy 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrIleLeuLeuAsnCysLeu 716
 Db 1966 ACTCTACTGGGAATTCGGGTGATCGTGAGGGTCTCCGGTGTCTGCTCAACGCTT 2025
 Qy 717 MetTyrIleMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
 Db 2026 ATGTACAAATGTGTACTACCGCTTTGGGAGGCTGTACACAGAACCAAGCGTCCACCA 2085
 Qy 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
 Db 2086 GGCITTTGACCTGTTCGAAATGCTGAGATTGGTAATAAGACTTTTGAGCTTGATGCTCG 2145
 Qy 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValIleAlaProAsp 776
 Db 2146 GAGGAAGCGTATACACAGAACACTGCTAGTCAGGATATACAAAGCTAAAGGACCTGGAT 2205
 Qy 777 AsnArg 778
 Db 2206 AATCGA 2211

RESULT 5

US-10-028-384-9
 ; Sequence 9, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian S1MP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 9
 ; LENGTH: 3094
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_008408
 ; DATABASE ENTRY DATE: 2000-11-01
 ; RELEVANT RESIDUES: (1)..(3094)
 US-10-028-384-9

Alignment Scores:

Pred. No.: 1,43e-224 Length: 3094
 Score: 2266.50 Matches: 435
 Percent Similarity: 70.73% Conservative: 104
 Best Local Similarity: 57.09% Mismatches: 71
 Query Match: 15.50% Indels: 8
 Gaps: 8

US-10-028-384-2 (1-826) x US-10-028-384-9 (1-3094)

Qy 61 GlyLeuSerGlnProAla---GlyTrpGln----- 69
 Db 7 GGGTTGAGTGGCGCTGAACGGATGCGGGGAGCAGAGTGGTTCCTGAGGAGCATC 66
 Qy 69 ----- 69
 Db 67 CGTGAGGTATTGTAATATCATCTGCTGCCACCCATTGATGCAAGATGACTAAGCTTGA 126
 Qy 70 -----SerLeuLeuSerPheThrIleLeuPheLeu 79
 Db 127 TTTTGGGATTGCTCTATGAGAGCAGGACACACTCTTAAAGCTTCTCATCTCTGTCGATG 186

Qy 80 AlaTrpLeuAlaClyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
 Db 187 GCTGCTGTGTATCTTTTCTACTCGCTTTTGTCTGTGTGAGATTGTAAGAGTGTCA 246
 Qy 100 HisGluPheAspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119
 Db 247 CATGAGTTTGTATCCGTACTTTAATTATCGGACTACCCGGTTTCTGGCTGAGAGGGTTT 306
 Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
 Db 307 TATAAATTCATAACTGGTTTGTATGACCGGGTGTGGTACCTTTTGGCGCAATCATTTGA 366
 Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
 Db 367 GGAACAATTTACCCAGTTTAATGATCACTTCTGCTGCAATCTACCATGTACTCATTTTC 426
 Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
 Db 427 TTCATATCATATTCGCAATTCGGAATGCTGTGTGTTCCTGGCCCACTTTTCTCTCT 486
 Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeu 199
 Db 487 TTCACCACCATCGTTACGTACCACCTTACCAAGAGCTCAAGGATGCGAGGCTGGGCTT 546
 Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
 Db 547 CTTGCTGCTGCCATGATGCTGTCTGTCTGGTATATTTCTCGATCTGTAGTGGCTTC 606
 Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIys 239
 Db 607 TATGATATGAGGAATTCATCTTTTCATGCTGCTTACTTACTACATGTCGATCAAG 666
 Qy 240 SerValIysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMet 259
 Db 667 GCAGTGAAGAGCTGTTCCTATCTATTGGGCTGCCAAGTGTGCCCTGCTATTCTACATG 726
 Qy 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 Db 727 GTCTCTTCAATGGGAGGCTATGTGTCTGTGATCAACTTGATTCCTCTACATGTCCTGTG 786
 Qy 280 LeuLeuLeuMetGlnArgTyrSerIysArgValTyrIleAlaTyrSerThrPheTyrIle 299
 Db 787 CTAATGCTGACAGCGCGTTTCTCACCGGATCTACGTAGCTACTGTACTGTATTACTGC 846
 Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 847 CTGGGACCATCTTTCTATGCAGATTTCTTTGTTGGTTTCCAGCGCTCTTTCATCA 906
 Qy 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 Db 907 GAACACATGCGAGCCTTTGGAGTGTGTGTCCTCTCAGATCCCATGCTTCTGTAGATTAC 966
 Qy 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 967 CTGGGAGCAGTGTGAATCCACAGCAATTCGAAGTCTTTCCGAGGTATCTCTCCCTG 1026
 Qy 360 AlaAlaGlyAlaValPheLeuSerVal---IleTyrLeuThrTyrThrGlyTyrIleAla 378
 Db 1027 ---GTTGGGTTTGTCT 1083
 Qy 379 ProTrpSerGlyValArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIlePro 398
 Db 1084 CCCTGGACAGCGGTTTCTACTCTCTGCTGGATCCCTTATGCTAAGAATAACATTTCC 1143
 Qy 399 IleIleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeu 418
 Db 1144 ATTATTGCAATCTGTTTCTGAGCAGCCAGCCCACTGCTCTCTCTCTCTCTCTCTCTCT 1203
 Qy 419 HisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAsp 438
 Db 1204 CAGCTCTTGTCTTCTATGTTTCCAGTTGGCTCTATTACTGCTTTTAGCAACCTGCTGAT 1263

NUMBER OF SEQ ID NOS: 176
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 92
 LENGTH: 3093
 TYPE: DNA
 ORGANISM: Mus musculus
 US-10-417-375-92

Alignment Scores:

Pred. No.: 1.43e-224 Length: 3093
 Score: 2266.50 Matches: 435
 Percent Similarity: 70.73% Conservative: 104
 Best Local Similarity: 57.09% Mismatches: 152
 Query Match: 51.50% Indels: 71
 DB: 18 Gaps: 6

US-10-028-384-2 (1-826) x US-10-417-375-92 (1-3093)

Qy 61 GlyLeuSerGlnProAla---GlyTrpGln----- 69
 Db 7 GGGTTGAGTGGCGCGCTGAACGGATGCGAGGGGAGCAGAGTGGTTCCTGAGGAGCATC 66
 Qy 69 ----- 69
 Db 67 CGTGAGGTATTGTGATATCATCATCTGCCACCCCATTCATGCTCAAGATGACTAAGCTTGA 126
 Qy 70 -----SerLeuLeuSerPheThrLeuPheLeu 79
 Db 127 TTTTTCGATGTCCTATGAGAACGAGCAGACACTTCTAAAGCTTCTCATCTCTGTCGATG 186
 Qy 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValleArgPheGluSerIleIle 99
 Db 187 GCTGCTGTGTATCTTTTCTACTCGCTTTTCTGCTGTGTGAGATTTGAAGTGTATC 246
 Qy 100 HisGluPheAspProTrpPheAsnTrpArgSerThrHisLeuAlaSerHisGlyPhe 119
 Db 247 CATGAGTTGTATCGGTACTTTAAATATCGGACTACCGGTTTCTGCTGAGGAGGCGTTT 306
 Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGly 139
 Db 307 TATAAATTCATACCTGTTTGTATGACCGGGTGTGACCTTTGGCCGAAATCATTGA 366
 Qy 140 GlyThrValTrpGlyLeuMetIleThrAlaGlyLeuIleHisTrpLeuAsnThr 159
 Db 367 GGAAACAATTTACCCAGGTTTAAATGATCATTCTGCTGCAATCTACCATGTTCTCCATTTC 426
 Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
 Db 427 TTCATATCATATGACATTGCGAATGTCGTGTTTCTGCCCCACCTTTTCTCTCT 486
 Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
 Db 487 TTCACCAACATCGTTATCGTACACCTTACCAAGAGCTCAAGGATCGAGGAGTGGGCTT 546
 Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySer 219
 Db 547 CTGCTGTCGCATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
 Qy 220 PheAspGlnGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeuTrpVallys 239
 Db 607 TATGATATGAAGGAATTTGCTATCTTTTGCATGCTGCTTACTTACTACATGTGGATCAAG 666
 Qy 240 SerValIleThrGlySerValPheThrThrMetCysCysLeuSerTrpPheTrpMet 259
 Db 667 GCAGTGAGAGATGTTTCTCATCTATGGCTGCCAGTGTGCTGCTGCTGCTGCTGCTGCTG 726
 Qy 260 ValSerAlaTrpGlyTrpValPheIleAlaLeuLeuIleAsnLeuLeuProLeuHisValPheVal 279
 Db 727 GTCTCTTCAGGGAGGCTGTGTTCTGATCAACTGATTTCTGATGCTGCTGCTGCTGCTGCTG 786
 Qy 280 LeuLeuLeuMetGlnAlaGlyTrpSerLysArgValTrpIleAlaTrpSerThrPheTrpIle 299
 Db 787 CTATGCTGACGAGCGGCTTTTCTACCGGATCTACGTAGCCTTACTGTTACTGCT 846

Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 847 CTGGGACCACTTCTTTCTATGAGATTCTCTTTGTTGTTTCCAGCCGCTCTTTCATCA 906
 Qy 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrp 339
 Db 907 GAACATGGCAGCCTTTGGAGTGTGTCTCTGTCAGATCCATGCTTCTGATGATAC 966
 Qy 340 LeuArgAspArgLeuThrLysGlnPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 967 CTGGCAGCAAGTGAATCCACAGCAATTCGAAGTCTTTTCCGGAGTGTATCTCCCTG 1026
 Qy 360 AlaAlaGlyAlaValPheLeuSerVal---IleTrpLeuThrTrpThrGlyTrpIleAla 378
 Db 1027'---GTTGGCTTTGCT 1083
 Qy 379 ProTrpSerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaIleHisIlePro 398
 Db 1084 CCTGGACAGGGGTTTCTACT 1143
 Qy 399 IleIleAlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeu 418
 Db 1144 ATTATGCACTCTTTCTGAGCACCCACCACTGCTCTCTCTCTCTCTCTCTCTCTCTCT 1203
 Qy 419 HisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAsp 438
 Db 1204 CAGTCT 1263
 Qy 439 GluArgValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetVal 458
 Db 1264 GCTCGGATTTTATCATCATGATGATGATGATGATGATGATGATGATGATGATGATG 1323
 Qy 459 ArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsn 478
 Db 1324 CGTCTAATGCTGATTTGTCACCTGTTATGTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 1383
 Qy 479 ValPheGluHisTrpLeuGlyAsp---AspMetLysArgGluAsnProProValGluAsp 497
 Db 1384 GTGCTGTCCATATATGAAAAATCTGGACATAGTGGCCCA----- 1425
 Qy 498 SerSerAspGluAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysVal 517
 Db 1426 -----GACAAGAGAGCAGCAGAG 1443
 Qy 518 ArgLysHisAlaThrGluGlnGluLysThrGluGlyLeuGlyProAsnIleLysSer 537
 Db 1444 CAACAGATTTCTACTTACCTTATTAAGATGAG-----GTGGCGAGT 1485
 Qy 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 Db 1486 GGGATGATATCTGTCATGCTTTTCTCTCATCATCATCATCATCATCATCATCATCATCAT 1545
 Qy 558 ThrSerAsnAlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThr 577
 Db 1546 ACCAGTGAAGCCTTTCT 1605
 Qy 578 ArgAsnIleLeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGlu 597
 Db 1606 AGGATCATTTTGTGATGCTCCAGAGCGTATTATGCTCGCTCCGTCACCAATCTCCAG 1665
 Qy 598 HisAlaArgValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMetAlaAsnArg 617
 Db 1666 GATGCAAAAGTCTATGCTGGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1725
 Qy 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
 Db 1726 ACAATTTTGTGACAAATAACACATGGATAATACCATATTTCTCTCGAGTAGGCGACGA 1785
 Qy 638 MetSerSerAsnGlnThrAlaAlaTrpLysIleMetArgThrLeuAspValAspTrpVal 657
 Db 1786 ATGGCATCCACAGAGAAAAAGCCTATGAATCATGAGGAGCTTGTATGTCAGTATG 1845

Db 398 CCGGTCATATTCGTGACATCTGGGTCTTCTGGCGCCGATCTTCAGTGGCTGACCTCC 457
Qy 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
Db 458 ATCTCCACCTACCTGCTGACCAAGAGCTGTGTGCGGGCGCGCTCTTCGCGGCC 517
Qy 203 CysPheLeuAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsp 222
Db 518 AGCTTCATCGCCATCGCGCTGGCTCATCAGTAGGTGGTGGCTGGATCGACGATAC 577
Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTrpLeuTrpValIleSerValIys 242
Db 578 GAGGCAATTCGATATTCGCCCTGCAGTTCACCTACTCTCTGGGTGGCTCAGTGAAG 637
Qy 243 ThrGlySerValPheThrTrpMetCysCysLeuSerTyrPheTyrMetValSerAla 262
Db 638 ACTGGATCCGGTCTCTGGTGGCGCGAGCCGCTTTGTCTTCTTACATGTGTGGCCG 697
Qy 263 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
Db 698 TGGGTGGCTACGTGTTTCATCATCACTGATACCCCTGCACGCTCTTCTGCTACTGCTCAT 757
Qy 293 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
Db 758 ATGGCAGTACTCGCGCGTCTGCTGACCGACTACAGACCTTCTACATCTCTGCTACTGCTCAT 817
Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
Db 818 CTGTTCTCCATGACAGATCCCTCTGGTGGATTCACACCGATACGACCATGCAACATG 877
Qy 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgasp 342
Db 878 GCTGGCTGGGAGTGTGTGCTCTTATGGCGGTGGCCACCTTGGCCATTTGACGCC 937
Qy 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGly 362
Db 938 GTGCTGTCCGCAACGAGTTCGGAGCTGTTCATCGTGGCGGATGTGTGGTGGCGTT 997
Qy 363 AlaValPheLeuSerValIleTyrLeuTyrThrGlyTyrIleAlaProTrpSerGly 382
Db 998 GCGGCTTTTGGCGCGTGGTGTCTCACCATGCTGGCGGTGGTGGCGCGGTGGAGTGA 1057
Qy 383 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 402
Db 1058 CGCTTCTACTGCTGTGGATCTGCTACGCGATCGCAAGATCCATCTCCATCTGATCC 1117
Qy 403 ValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHisIleLeuVal 422
Db 1118 GTGTGGAGCATACAGCCACCATTTGTTCTGTTCTTCTTGTATCTGCACATCTGGTG 1177
Qy 423 CysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPhe 442
Db 1178 TGGCGCTTCCAGTGGGAGTGTGTACTGATCAAGCAGATCAACGACGAGCGGTTTC 1237
Qy 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
Db 1238 GTGGTGTGTACGCGCATCAGTGGCGTTTCTTCGCTGGTGTGTGTGTGTGTGTGTGTGT 1297
Qy 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
Db 1298 ACCCTCAGCGCGT 1357
Qy 483 TyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSerSerAspGlu 501
Db 1358 TTCCTGCAAGAGATTCGCTTAACGATGGGCACAGCCTAAGCCGACCCACCGAGTGT 1417
Qy 502 Asp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArg 518
Db 1418 GATGAAGCTGAGGATTCATTCAGAGAAGACGCTGTACGACAGGCTGGCAAGCTG--- 1474
Qy 519 LysHisAlaThrGluGlnGlyThrGluGlu---GlyLeuGlyProAsnIleLysSer 537

Db 1475 AAGCATCGTACTAAGCATGATGCCCCAGCAGGATAGTGGCGTCACTCCAACTGAGAGT 1534
Qy 538 IleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
Db 1535 ATTGTATTATTGGCGGTTCATATGCTGTGTATGATGTTCGCTGCTCCACTGACCTGGGTG 1594
Qy 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
Db 1595 ACCAGGAAATGCTCTACTCCAGTCCCTCCATGTCTTGGTCTTCCACACAGTCAAGATGGA 1654
Qy 577 ThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAsp 596
Db 1655 TCCCGCAACATTTAGACGATTCAGAGAGGCTTACTACTGGCTTCCGAGAACACTGCC 1714
Qy 597 GluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 616
Db 1715 GATGATGCTCGGTATGTCTTGGTGGATTAAGGATACAGATAGCGGGAATGCGAAC 1774
Qy 617 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 636
Db 1775 AGAACGACGCTAGTGGATAATAATACGTGGAACAATAGTACATAGCGCTGGTGGCAAG 1834
Qy 637 AlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyr 656
Db 1835 GCATGTCTTCAACGAGGAGAGTCTTACGNAATTATGACATCTCTTGAGTGGACTAC 1894
Qy 657 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
Db 1895 GTTTTGGTGTATCTTTGGCGTGTGATCGCTATTCTGCGGATGATATCAACAAGTTCCTG 1954
Qy 677 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
Db 1955 TGGATGCTCGAATGTCTGAGGAGAGCATCCCAAGACATTAAGAAAGCGATTACTTT 2014
Qy 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716
Db 2015 ACCGACGGGGTGAATTCAGGGTAGATCCGAAGGTGCTCCGGCCCTGCTCAACTGCCCT 2074
Qy 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
Db 2075 ATGTACAAATTAAGTACTACAGATTCGGGNAATTGAAGTTGGACTACAGAGTTCATCT 2134
Qy 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
Db 2135 GGATATGATCGCACACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG 2194
Qy 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp 776
Db 2195 GAGAGGCTTACACACAGACACTGGCTTGTTCGATCTATAGGTGAGAGACCGCAT 2254
Qy 777 -----AsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGln 794
Db 2255 GAGTTCAATAGACCACTCACTGAAGACCAAGAGAGAGACG-----ATTCTCCAGCA 2305
Qy 795 LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814
Db 2306 AACTTCATTTCCAGAAAGAACTCTAAGCGTCCAGAGGCTACATACGAACCCGCGTT 2365
Qy 815 PheLysGlyGlyLysLysLysSerLysThr 825
Db 2366 GTTGTAAAGGAAACGAACTTTGAATAAAC 2398

RESULT 4

US-10-417-375-92
; Sequence 92, Application US/10417375
; Publication No. US2004021928A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15

Db 1205 TGGAGTGGCAGGTTTATCTACTATGGATGGATCTGGGTATCCAAATATACATCTCCATT 1264
Qy 400 lleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHis 419
Db 1265 ATTGCATCAGTGTCTGAACATCCAGCTACGACATGGGTGTCTTCTTCTTTGATCTACAT 1324
Qy 420 lleLeuValCysThrPheProAlaGlyLeuTrpPheCysAlaIleLysAsnIleAsnAspGlu 439
Db 1325 ATTCTTGTATGACTCTCCAGCAGGCTATGGTTCGCATCAAAATATCAACGATGA 1384
Qy 440 ArgValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArg 459
Db 1385 AGAGTATTGTGGTCTGTATGCGCATCAGTCTGTGTACTTTCGCGAGTGTGGTGGG 1444
Qy 460 LeuMetLeuThrLeuProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
Db 1445 CTGATGCTGACTCTGACCCCGTGTCTGCGATGCTGCGCCATCGCTTCTCCATGTT 1504
Qy 480 PheGluHisTrpLeuGlyAspAspMetIleArgGluAsnProProValGluAspSerSer 499
Db 1505 TTTGAGCACTATTGGGGGATGACATGAAAGGGAACCCACCTGTGGAGGACAGCAT 1564
Qy 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArgLys 519
Db 1565 GATCAGGATGACAAAGAAACCCAGGAACTGTATGACAAGCAGGTAAGTGAAGAG 1624
Qy 520 HisAlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
Db 1625 CATGTGACAGAGCAAGAGAACCTGGAAGAGGCTTTGGGCCCCAACATCAAAAGCATGTG 1684
Qy 540 ThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpValThrSer 559
Db 1685 ACCATGCTGATGCTCATGCTCTGATGATGTTGCGGTCTCCATGCACTGGGTGCAACAGC 1744
Qy 560 AsnAlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsn 579
Db 1745 AACGCTACTCCAGTCCAGGTGTGGTCTCTGCTTCTTACATCATGATGATGATGATGAT 1804
Qy 580 lleLeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAla 599
Db 1805 ATATTAGATGATTTAGAGAAGCGTACTTTTGGCTGAGACAAACACGATGAACAGCC 1864
Qy 600 ArgValMetSerTrpTrpAspTrpGlyGlnIleAlaGlyMetAlaAsnArgThrThr 619
Db 1865 CGGGTCATGTCGTGTGGGACTACGGCTATCAGATTCGTGGCATGGCCCAACAGGACCAT 1924
Qy 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
Db 1925 CTGGTGGATAACAACACCTGGAACAACAGCCACATGCTGCTGCGAAGCTATGCTCT 1984
Qy 640 SerAsnGluThrAlaAlaTrpLysIleMetArgThrLeuAspValAspTrpValLeuVal 659
Db 1985 TCCAATGAACGGCCGCTATAAATCATGAGGTCCCTTATGTCGATGATGATGATGATGAT 2044
Qy 660 llePheGlyGlyValIleGlyTrpSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
Db 2045 ATTTCGAGGAGTGATTTGGCTATTCCGGGACGATATCAACAAGTTCTCTGGATGTC 2104
Qy 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTrpPheThrProGln 699
Db 2105 AGGATAGCTGAAGGGAGATCCCAAGACATCCGGAAGGTGACATTTTCCACCCAGCAG 2164
Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTrpLys 719
Db 2165 GGAGAGTTCCGAGTAGACAAAGCTGGTCTCTACTCTGTATAAATCGCTTATGATATAA 2224
Qy 720 MetSerTrpTrpArgPheGlyGluMetGluLeuAspPheArgThrProProGlyPheAsp 739
Db 2225 ATGTCATACTACGATTTGGAGAATGCGATGATTTTCCGCTCTCCCGAGGCTTTGAC 2284
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759

Db 2285 CGAACACGTAACTGCTGAGATTGGAATAAAGACATTAAATTCACGATTTGGAGAACT 2344
Qy 760 PheThrSerGluHisTrpLeuValArgIleTrpLysValLysAlaProAspAsnArgGlu 779
Db 2345 TTTACATCAGAGACCTGGCTTGTGAGGATATATAAAGTGAAGACCTGTGACACAGGAG 2404
Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTrpLeuSerLys 799
Db 2405 ACATAGGTCAACAACCTCGAGTCACCAACATCGTCCCAACACAGAGTAITTTGTCAAG 2464
Qy 800 LysThrThrLysArgLysArgGlyTrpIleLysAsnLysLeuValPheLysGlyLys 819
Db 2465 AAGACTACTAAAGGAGCGTGGCTACGTAAATAAAGTAAAGTGTGTTTAAAGAAAGCAAG 2524
Qy 820 LysIleSerLysLysThrVal 826
Db 2525 AAGACCTCTAAGAAGACTGTT 2545

RESULT 3
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Kevin
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

Alignment Scores:
Pred. No.: 7,3e-236 Length: 2417
Score: 2950.00 Matches: 557
Percent Similarity: 83.79% Conservative: 89
Best Local Similarity: 72.24% Mismatches: 113
Query Match: 67.03% Indels: 12
DB: 15 Gaps: 7

US-10-028-384-2 (1-826) x US-10-028-384-7 (1-2417)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
Db 98 AGCAAGGTGGCTGGCTACAGAGCCCTAATCACTCGCCATCTGCTAATCGCTGGGTG 157
Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
Db 158 GCGGATTTCTCTCGCTCTTCGCGCTATCCGTTTCGATCGCTATTCATGATGTT 217
Qy 103 AspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTrpGluPhe 122
Db 218 GATCCGTGGTTCAACTACCGGCGCCACCGCTCATCATGGTGGTGGTGGTGGTGGT 277
Qy 123 LeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyValGlyValGlyVal 142
Db 278 CTCAACTGGTTCGACGAGCGCATGGTATCGCTCGCAGGATTTGTGGCGGTACCGTC 337
Qy 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
Db 338 TATCCCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 397
Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182

Db 1981 TTTGAGGGGTTATGGCTATCTGGTGATGATCAACAAATTTCTCTGGATGGTAGG 2040
 Qy 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 2041 ATAGCTGAAGAGAACATCCCAAGACATTCGGGAAGTGAATTTTACCCACAGGGA 2100
 Qy 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 2101 GAAATCCGTGTAGACAAAGCAGGATCCCTACATTTGTGTGAATTCCTTATGATAAAATG 2160
 Qy 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlnGlyPheAspArg 740
 Db 2161 TCATCTACAGATTTGGAGAAATGCAGCTGATTTTCGTACACCCACAGGTTTACCGA 2220
 Qy 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
 Db 2221 ACAGCTAATGCTGAGATTTGGAATAAGGACATTAATAATTCAAACATTTGGAAGAAGCCATT 2280
 Qy 761 ThrSerGluHisTyrPheValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2281 ACATCAGAACACTGGCTGTTTGGATATATAAAGTAAAGCCCTGATTAACAGGAGACA 2340
 Qy 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAAATATTGTCAAAGAAG 2400
 Qy 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 Db 2401 ACTACCAAGAGAGCGTGGCTACATTAATAAATAAGCTGTTTTTAAGAAGAGCAAGAAA 2460
 Qy 821 IleSerLysLysThrVal 826
 Db 2461 ATATCTAAGAAGACTGTT 2478

RESULT 2

US-10-028-384-3
 ; Sequence 3, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 2710
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: AK018758
 ; DATABASE ENTRY DATE: 2001-07-05
 ; RELEVANT RESIDUES: (1)..(2469)
 US-10-028-384-3

Alignment Scores:
 Pred. No.: 0 Length: 2710
 Score: 4263.50 Matches: 805
 Percent Similarity: 77.70% Conservative: 3
 Best local Similarity: 97.34% Mismatches: 14
 Query Match: 96.88% Indels: 5
 DB: 15 Gaps: 2

US-10-028-384-2 (1-826) x US-10-028-384-3 (1-2710)

Qy 1 MetaAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
 Db 77 ATGGCGGAGCCCTCGGCCCCGGAGAGCAACAGACGTCCTCAACTCGTCCCGCGTGG 136

Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
 Db 137 AGCGGCTCATGGCTCTGGGAAACAGCCGCCACCGGACCATGGCGCCGGAACCCAGAGC 196
 Qy 41 AlaHisLysAlaAlaGlyAlaAlaProProLysProAla---ProAlaGlyLysSer 59
 Db 197 GCCTCCAGCGCGCGCGG-----CCGAAGCGCGGCGCCCGCGGCGCTGTCC 244
 Qy 60 GlyGlyLeuSerGlnProAlaGlyTyrGlnSerLeuSerPheThrIleLeuPheLeu 79
 Db 245 GGGGGCTTGTGACAGCGCGCGGTGGCAGTCTGTCTCTCTCTTCCATCTCTCTCTCT 304
 Qy 80 AlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIle 99
 Db 305 GCCTGCTGCCCGCTTCAGCTCGCGCTCTTGGCGGTCATCCCTTCGAGAGCATCATC 364
 Qy 100 HisGluPheAspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119
 Db 365 CACGAGTTCGACCCGCTGGTTAACTATAGATCAACACATCATCTTGTGATCTCATGATTC 424
 Qy 120 TyrGluPheLeuAsnTyrPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGly 139
 Db 425 TATGATTTCTAAATTTGGTTTGTGAAAGAGCATGGTACCCTCGGAGAGATAGTGGGT 484
 Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleThrIleLeuAsnThr 159
 Db 485 GGCACCGTTTACCAGGGTTGATGATAACAGCTGGCTTATTCATTGGATTTTAAATACA 544
 Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
 Db 545 TTGAACATACAGTTTCAATAAGAGATGTGTGTTTCTTGTGACCAACTTTTAGCGGC 604
 Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsnGlnGlyAlaGlyLeu 199
 Db 605 CTTTACATCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAACCAAGAGAGAGACT 664
 Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
 Db 665 CTAGCTGCTGTTCTTCAATGCTATCGTACAGGATCAATCTCGGTGAGTGGCGGATTC 724
 Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLys 239
 Db 725 TTTGATAATGAAGGCAATGCCATTTTTCGGCTTCACTTCACTTCACTTCACTTCACTT 784
 Qy 240 SerValLysThrCysSerValPheThrMetCysCysCysLeuSerTyrPheTyrMet 259
 Db 785 TCTGTGAAGACCGGGTCTGTGTCTGGACAAATGTCTGCTCTGTCTGTCTGTCTGTCT 844
 Qy 260 ValSerAlaTyrGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 Db 845 GTCTCTGCGTGGGAGGTTATGTTTCACTCACTCACTCACTCACTCACTCACTCACTCA 904
 Qy 280 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
 Db 905 TTCTGCTGATGACAGGATGACAGAGAGTCTACATAGCATATAGCATTTGTACATT 964
 Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 965 GTGGGTTTAAATATATCCATGACAGATACCTTTTGTGGGATTTGAGCCCAATCAGAACA 1024
 Qy 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 Db 1025 GAGCATGCGAGCTGCGAGTGTCTTGGCTGTGCGTGTGCGAAGTTTACGCTTTTTCAGTAT 1084
 Qy 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 1085 CTGAGAGACCGGTTGACAAAACAGGAGTTCAGACCCCTTTTCTTTTGGGTGTCTCACTA 1144
 Qy 360 AlaAlaGlyValAlaPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
 Db 1145 GCTGCGAGGCGTGTGTTTCTTGTGTCTATCTATCTGACATACACAGGTTATATTGACCA 1204
 Qy 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399

OTHER INFORMATION:

US-10-028-384-1

Alignment Scores:

Pred. No.: 0
 Score: 4401.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 15
 Gaps: 0

US-10-028-384-2 (1-826) x US-10-028-384-1 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerHisHisSerSerLeuAsnSerSerProTrp 20
 DB 1 ATGGCGGAGCCCTCGCGCCCGGAGAGCAAGCAAGTCTCGTCCCTCAACTCGTCCCGTGG 60
 QY 21 SerGlyLeuMetAlaLeuGlyVAnSerArgHisGlyHisGlyProGlyValAlaGlnCys 40
 DB 61 AGTGGCCCTCATGGCCCTGGGNAACAGCCGCGCAAGCCACCGGGCCCGGGCCCGAGTGC 120
 QY 41 AlaHisIysAlaAlaGlyAlaAlaProProGlyProAlaProAlaGlyLeuSerGly 60
 DB 121 GCGCACAGGCGGCGGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGGCTGTCGACCGCGGCTGGGTGGCAGTCCGCTCTCTCTTCCACCATCTCTCTCTGGCC 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
 DB 241 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCGCTTCGAAAGCATCATCCAC 300
 QY 101 GluPheAspProTrpPheAsnTrpArgSerThrHisLeuAlaSerHisGlyPheTrp 120
 DB 301 GAGTTCACCGCGTGGTTAACTATAGATCAACACATCACTTCGCTCATCGGTTCAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATGGTTTGTATGAAGAGCATGTTATCCATAGAGAAAGATAGTAGTGTGT 420
 QY 141 ThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
 DB 421 ACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTCATTTGGATTTTAAATACATGG 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 481 AACATAACTGTTCATATAAGAGACGTATGTGTCTTGCACCAACTTTTAGCGGCCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuLeu 200
 DB 541 ACATCTATATCTACTTTCTCTTACAGAGAACTTTGGAAACAGGACAGGACATTTTA 600
 QY 201 AlalaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPhe 220
 DB 601 GCTGCTTGTATTTATTTGTTATTTACAGGCTACATATCTCGGTGAGTCTGGATCCCTTT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeuTrpValIysSer 240
 DB 661 GATTAATGAAGCATTTGCTATTTTGGACCTTTCAGTTTACATATTTTATTTTGGTTAAATCT 720
 QY 241 ValIysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetVal 260
 DB 721 GTAAAAAATCGGGTCAGTCTTTTGGACAAATGTGCTGCTTATCTTATTTATATATGTC 780
 QY 261 SerAlaTrpGlyGlyTrpValPheIleLeuAsnLeuIleProLeuHisValPheValLeu 280
 DB 781 TCTGCTGGGTGGTGTATGATTTATCATCAATCTTATTTCCATGCAATGATTTTGTGTTG 840
 QY 281 LeuLeuMetGlnArgTrpSerLysArgValTrpIleAlaTrpSerThrPheTrpIleVal 300
 DB 841 TTACTGTAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGTG 900

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTTTAAATATATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeu 340
 DB 961 CACATGCGAGCTCGAGGTGTCTTTGCAATTCGTCGAGCTTATGCTTTCTTCAGTATCTG 1020
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGACCGATTAAACAAACACAGAGTCCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValIleTrpLeuThrTrpGlyTrpIleAlaProTrp 380
 DB 1081 GCAGGTGCTGTCTCTTATGTCATCTATTTGACTTATACAGGTACATTTGACCATGG 1140
 QY 381 SerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleIle 400
 DB 1141 AGTGGCAGGTTTATTTTCAATTTGGGATCTGGGTATGCAAAATACACATTTCAATTTAT 1200
 QY 401 AlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheAspLeuHisIle 420
 DB 1201 GCATCATGTCGTGAGCATCAACCTAGACTTGGGTGTCTTTCTTCTTGTGATCTACATATT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLeuLysAsnIleAsnAspGluArg 440
 DB 1261 CTGTATGTACCTTCCACAGAGCCCTTTGGTCTGCATCAAAAATATCAACGATGAAGA 1320
 QY 441 ValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTTGTTGCTCTATATCAATCACTGCTGTCTTCTTGTGAGTGTGTCGAGCTG 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 DB 1381 ATGTTGACTTTGACTCCAGTCTGTGTATGCTGTCTGCAATTCCTTTTCAAATGTTT 1440
 QY 481 GluHisTrpLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCACCATTATTTGGGGATGACATGAAGAGGAAAAATCCACCTGTGGAGACAGCATGTAT 1500
 QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArgLysHis 520
 DB 1501 GAGGTATGACAAAGAAACCAAGCAATTTGTATGATAGGCGAGTTAAGTGGAGAAACAT 1560
 QY 521 AlaThrGluGlnGlyThrGluGluGlyLeuGlyProAsnIleLysSerIleValThr 540
 DB 1561 GCAACTCAACAGCAAAACCTGAGAGGATTTAGGCCCTTAATATAAAGCATTTGTCACC 1620
 QY 541 MetLeuMetLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 DB 1621 ATGTTGATGCTGATGTATTTGATGATGTTGCTGCTCCACTGTACCTGGGTCAACAGCAT 1680
 QY 561 AlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsnIle 580
 DB 1681 GCCTACTCTAGTCCAAAGTAGTCTGCTGCTCATCAATCATGATGGACCCAGCAATATC 1740
 QY 581 LeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 DB 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTTAGGCAAAATACAGATGAACATGCACGA 1800
 QY 601 ValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 DB 1801 GTAAATGCTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860
 QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 DB 1861 GTGGATTAATAACCTGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATGTCTTCT 1920
 QY 641 AsnGluThrAlaAlaTrpLysIleMetArgThrLeuAspValAspTrpValLeuValIle 660
 DB 1921 AATGAACACAGCAGCTTATAAATCATAGGACTCTAGATGTAGATTATGTTTGTGTTAT 1980
 QY 661 PheGlyGlyValIleGlyTrpSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 15, 2004, 06:02:06 ; Search time 1039.95 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 4401
Sequence: 1 MAEPAPESKHKSLNSSPW.....GVYKXKLVFKGKISKXTV 826

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131523_9140/app_query.fasta_1.3740
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DOCLIGN=200 -THR SCORE=pcpt -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384@cgn_1_1_2577@runat_14122004_131523_9140
-NCPU=6 -ICPU=3 -NO_WAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=130 -WARN TIMEOUT=130 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodaa/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodaa/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodaa/1/pubpna/US06_PUBCOMB.seq.*
4: /cgn2_6/ptodaa/1/pubpna/US07_NEW_PUB.seq.*
5: /cgn2_6/ptodaa/1/pubpna/US07_PUBCOMB.seq.*
6: /cgn2_6/ptodaa/1/pubpna/US08_PUBCOMB.seq.*
7: /cgn2_6/ptodaa/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodaa/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodaa/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodaa/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodaa/1/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodaa/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodaa/1/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodaa/1/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodaa/1/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/ptodaa/1/pubpna/US10_PUBCOMB.seq.*
17: /cgn2_6/ptodaa/1/pubpna/US10_PUBCOMB.seq.*
18: /cgn2_6/ptodaa/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodaa/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodaa/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodaa/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 4401 | 100.0 | 2481 | 15 | US-10-028-384-1 |
| 2 | 4263.5 | 96.9 | 2710 | 15 | US-10-028-384-3 |
| 3 | 2950 | 67.0 | 2417 | 15 | US-10-028-384-7 |
| 4 | 2266.5 | 51.5 | 3093 | 18 | US-10-417-375-92 |
| 5 | 2266.5 | 51.5 | 3094 | 15 | US-10-028-384-9 |
| 6 | 2262 | 51.4 | 3046 | 18 | US-10-417-375-95 |
| 7 | 2250 | 51.1 | 2472 | 15 | US-10-171-581-112 |
| 8 | 2250 | 51.1 | 2472 | 15 | US-10-028-384-11 |
| 9 | 2250 | 51.1 | 2472 | 15 | US-10-172-118-742 |
| 10 | 2250 | 51.1 | 2472 | 16 | US-10-342-887-742 |
| 11 | 2241 | 50.9 | 5404 | 18 | US-10-417-375-99 |
| 12 | 2241 | 50.9 | 5407 | 18 | US-10-417-375-97 |
| 13 | 2181.5 | 49.6 | 2466 | 16 | US-10-320-797-2305 |
| 14 | 2179.5 | 49.5 | 2232 | 15 | US-10-128-714-7139 |
| 15 | 2164 | 49.2 | 2256 | 15 | US-10-032-585-6323 |
| 16 | 2120.5 | 48.2 | 2779 | 17 | US-10-437-963-99904 |
| 17 | 2118.5 | 48.1 | 2839 | 18 | US-10-425-115-150745 |
| 18 | 2117 | 48.1 | 2681 | 16 | US-10-425-114-14408 |
| 19 | 2023 | 46.0 | 2603 | 15 | US-10-128-714-6139 |
| 20 | 2023 | 46.0 | 2969 | 15 | US-10-128-714-139 |
| 21 | 2023 | 46.0 | 4603 | 15 | US-10-128-714-5139 |
| 22 | 2021.5 | 45.9 | 2157 | 9 | US-09-801-368-387 |
| 23 | 2021.5 | 45.9 | 2157 | 18 | US-10-793-639-318 |
| 24 | 2021.5 | 45.9 | 2733 | 15 | US-10-028-384-5 |
| 25 | 1958 | 44.5 | 1828 | 10 | US-09-945-527-62 |
| 26 | 1941.5 | 44.1 | 4738 | 16 | US-10-320-797-305 |
| 27 | 1939.5 | 44.1 | 2882 | 15 | US-10-320-797-1305 |
| 28 | 1897 | 43.1 | 1848 | 15 | US-10-128-714-2139 |
| 29 | 1896.5 | 43.1 | 1343 | 10 | US-09-974-879-133 |
| 30 | 1896.5 | 43.1 | 1543 | 10 | US-09-305-736-133 |
| 31 | 1896.5 | 43.1 | 1543 | 10 | US-09-818-683-133 |
| 32 | 1896.5 | 43.1 | 1543 | 15 | US-09-818-683-133 |
| 33 | 1896.5 | 43.1 | 1543 | 16 | US-10-621-401-133 |
| 34 | 1894.5 | 43.0 | 1209 | 15 | US-10-106-698-330 |
| 35 | 1894.5 | 43.0 | 1209 | 16 | US-10-264-237-412 |
| 36 | 1882.5 | 42.8 | 3141 | 18 | US-10-425-115-130787 |
| 37 | 1855 | 42.1 | 1369 | 15 | US-10-128-714-1139 |
| 38 | 1851.5 | 42.1 | 3197 | 16 | US-10-424-599-111541 |
| 39 | 1610.5 | 36.6 | 2244 | 17 | US-10-437-963-48342 |
| 40 | 1581 | 35.9 | 2660 | 16 | US-10-264-049-630 |
| 41 | 1552.5 | 35.3 | 1114 | 16 | US-10-296-115-629 |
| 42 | 1288.5 | 29.3 | 1728 | 16 | US-10-424-599-122476 |
| 43 | 975.5 | 22.2 | 1094 | 17 | US-10-437-963-99902 |
| 44 | 871 | 19.8 | 500 | 9 | US-09-998-588-1643 |
| 45 | 763 | 17.8 | 485 | 10 | US-09-918-955-11283 |

ALIGNMENTS

RESULT 1
US-10-028-384-1
Sequence 1, Application US/10028384
Publication No. US20030148285A1
GENERAL INFORMATION:
APPLICANT: COMPATIGENE INC.
APPLICANT: PERREAU, Claude
APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028,384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 2481
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2481)

; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3090
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3090

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-----|
| Pred. No.: | 2,03e-50 | Length: | 594 |
| Score: | 544.00 | Matches: | 102 |
| Percent Similarity: | 78.31% | Conservative: | 28 |
| Best Local Similarity: | 61.45% | Mismatches: | 32 |
| Query Match: | 12.36% | Indels: | 4 |
| DB: | 4 | Gaps: | 3 |

US-10-028-384-2 (1-826) x US-09-248-796A-3090 (1-594)

| | | | |
|----|-----|--|-----|
| QY | 76 | IleLeuPheLeuAlaTrpLeuAlaGly-----PheSerSerArgLeuPheAlaValIle | 93 |
| DB | 100 | ATTATATTATATCG--ATAGCAGGTGCAGCTATTCTCTCGTTTATTTCCGTGATT | 156 |
| QY | 94 | ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgSerThrHisHis | 113 |
| DB | 157 | CGATTTGAAGATATATTCATGAAATCGATCCCTGGTTCATTTCCGACACCAATAT | 216 |
| QY | 114 | LeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro | 133 |
| DB | 217 | TTAGTCACTCATTCCTTTATGAATTTTGAATGGTTTGATGATAGAACTTGGTACCCA | 276 |
| QY | 134 | LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIle | 153 |
| DB | 277 | TGGGAAGAGTCACTGGTGGTACTTATATCCCGTTTATGGTGACTTCAGGTGCCATT | 336 |
| QY | 154 | ---HisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPhe | 172 |
| DB | 337 | TGGCATATTTCACGTGATTGGTTGGCTTACCCGTTGATATTATAAATATTGTGTTTA | 396 |
| QY | 173 | LeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeu | 192 |
| DB | 397 | TTAGCACCATTTTCTCGGATTAACTCGAATTGCAATTGACTTATTTTGTACTAAGAAATG | 456 |
| QY | 193 | TrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIle | 212 |
| DB | 457 | AACGATTCCTAGTCACGATTATTGGCAGCTATATTATGGGATTGCCCCAGGTTATATT | 516 |
| QY | 213 | SerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPhe | 232 |
| DB | 517 | TCAAGATCAGTGGCTGGTCTTATGATPATGAAGCAATTGCCAATTCTTTATTAATGGCA | 576 |
| QY | 233 | ThrTyrTyrLeuTrpVal | 238 |
| DB | 577 | ACATTTTATTTCTGGATT | 594 |

Search completed: December 15, 2004, 11:58:29
Job time : 200.664 secs

QY 574 HisAspGlyThrArg 578
 |||||
 Db 21 CAAGATGGATCCCGC 7

RESULT 13

US-09-270-767-16600/c
 ; Sequence 16600, Application US/09270767
 ; Patent No. 6703491

GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.

FILE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094

CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17

NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 16600

LENGTH: 560

TYPE: DNA

ORGANISM: *Drosophila melanogaster*

US-09-270-767-16600

Alignment Scores:
 Pred. No.: 3,68e-56 Length: 560
 Score: 595.50 Matches: 119
 Percent Similarity: 78.38% Conservative: 26
 Best Local Similarity: 64.32% Mismatches: 33
 Query Match: 13.53% Indels: 7
 Db: 4 Gaps: 5

US-10-028-384-2 (1-826) x US-09-270-767-16600 (1-560)

QY 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHis 419
 |||||
 Db 558 ATTCATCCGTCGGAGCATGACCCACCATGGTTCTCGTCTCTTCATCTGCAC 499
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysAlaIleAsnIleAsnGlu 439
 |||||
 Db 498 ATCATGTGCGCGCTCCAGTGGGGTGTGTACTGTCATCAAGCAGATCAACACGAG 439
 QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 |||||
 Db 438 CGCGTTTCGTGGTCTGTACGCCATCAGTCGGGTTTACTTCGCTGGTGTGTGCGT 379
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 |||||
 Db 378 TTGATGTGACCTCCAGCGGGTGGTGTGCATCTGGCGGAGTGGCCCTTTTCGGGACTG 319
 QY 480 PheGluHisTyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSer 498
 |||||
 Db 318 TTGCATGTGTCCTGCAAGAGGATTCGTCTAACCGAATGGCCAGCAGCCATAGCCAGCC 259
 QY 499 SerAspGluAsp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGly 515
 |||||
 Db 258 ACCAAGTGGATGAAGCTGAGGATTCATTGAGAAGAACAGCCGTGTACACAAAGCTGGC 199
 QY 516 LysValArgLysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsn 534
 |||||
 Db 198 AAGCTG---AAGCATGCTACTAAGCATGATGCCAGCAGSACTATGGCGTCAAGTCCCAAC 142
 QY 535 IleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCys 554
 |||||
 Db 141 CTGAAGAGATATGTATTTTGGCGGTCTAATGCTGTATGATGTTCGCTGTCCACTGC 82
 QY 555 ThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn--- 573
 |||||
 Db 81 ACGTGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATGCTCTTGGCTTTCCACACAGT 22
 QY 574 HisAspGlyThrArg 578
 |||||
 Db 21 CAAGATGGATCCCGC 7

RESULT 14

US-09-270-767-27262
 ; Sequence 27262, Application US/09270767
 ; Patent No. 6703491

GENERAL INFORMATION:

APPLICANT: Homburger et al.

FILE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094

CURRENT APPLICATION NUMBER: US/09/270,767

CURRENT FILING DATE: 1999-03-17

NUMBER OF SEQ ID NOS: 62517

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 27262

LENGTH: 503

TYPE: DNA

ORGANISM: *Drosophila melanogaster*

US-09-270-767-27262

Alignment Scores:
 Pred. No.: 1.53e-51 Length: 503
 Score: 553.00 Matches: 105
 Percent Similarity: 75.30% Conservative: 20
 Best Local Similarity: 63.25% Mismatches: 41
 Query Match: 12.57% Indels: 0
 Db: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-27262 (1-503)

QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 |||||
 Db 2 GGGTATATCTTCGATCCGTGGCGGATCGTACGACATGAGGATCGCATTTCTGCG 61
 QY 230 LeuGlnPheThrTyrTyrIleTrpValLysSerValLysThrGlySerValPheThr 249
 |||||
 Db 62 ATGCTCTTCCACTACTATTTGTGGATCAAGCGGTAAAGACGGSCAGATCTTTTGTGCG 121
 QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
 |||||
 Db 122 GCTATGTCGGCATTTGGCTACTTCTATATGTTCTCTCGTGGGGTGGCTATGTTCTG 181
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
 |||||
 Db 192 ATTAACCTAATCCCGTCGACGTCGTGGCGGTGATGATCACCGGACGTTTCTGCAAGG 241
 QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuLeuSerMetGlnIlePro 309
 |||||
 Db 242 ATCTACATAGCATACAGCAGCTATCTGCTGGCACCATTCTGTCGATGCAGATCTCG 301
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 |||||
 Db 302 TTGTGGGATTCGAACCCATCCAGAGCTCCGAACACATGCTGGCAGCTGGGAACCTTTGGC 361
 QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
 |||||
 Db 362 CTGTGCAGATTCAGCTTTTGTCCACTATCTGGCTCGCAGATCCAGGATCACTTC 421
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValle 369
 |||||
 Db 422 GATCTCTCTTCAAGACAGTGGTTTCCAGTGTGTTTACTGTGGTGTGTTCTGCTGGGTACC 481
 QY 370 TyrLeuThrTyrThrGly 375
 |||||
 Db 482 CTGCTCACGCTTACCGG 499

RESULT 15

US-09-248-796A-3090

; Sequence 3090, Application US/09248796A

; Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al

FILE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: 107196.132

CURRENT APPLICATION NUMBER: US/09/248,796A

RESULT 11

US-09-328-111-332
 ; Sequence 332, Application US/09328111
 ; Patent No. 6262333
 ; GENERAL INFORMATION:
 ; APPLICANT: Endege, Wilson O.
 ; APPLICANT: Steirmann, Kathleen E.
 ; APPLICANT: Astle, Jon H.
 ; APPLICANT: Burgess, Christopher C.
 ; APPLICANT: Bushnell, Steven E.
 ; APPLICANT: Carroll III, Eddie
 ; APPLICANT: Catino, Theodore J.
 ; APPLICANT: Derti, Adnan
 ; APPLICANT: Ford, Donna M.
 ; APPLICANT: Lewis, Marcia E.
 ; APPLICANT: Monahan, John E.
 ; APPLICANT: Schlegel, Robert
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 ; TITLE OF INVENTION: PRODUCTS
 ; FILE REFERENCE: CDD-257 (US)
 ; CURRENT APPLICATION NUMBER: US/09/328.111
 ; CURRENT FILING DATE: 1999-06-08
 ; EARLIER APPLICATION NUMBER: US 60/088,801
 ; EARLIER FILING DATE: 1998-06-10
 ; NUMBER OF SEQ ID NOS: 850
 ; SOFTWARE: PastSeq for Windows Version 3.0
 ; SEQ ID NO 332
 ; LENGTH: 616
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1) -- (616)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-328-111-332

Alignment Scores:

Pred. No.: 4,33e-68 Length: 616
 Score: 704.00 Matches: 148
 Percent Similarity: 86.29% Conservative: 3
 Best Local Similarity: 84.57% Mismatches: 17
 Query Match: 16.00% Indels: 8
 DB: 3 Gaps: 2

US-10-028-384-2 (1-826) x US-09-328-111-332 (1-616)

Qy 208 ValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIle 227
 Db 5 GTACCAGCTACATATCTCGTCACTAGCTGGATCTTTCATATGAAAGCAATTCCTATT 64
 Qy 228 PheAlaLeuGlnPheThrTyrTyrLeuTyrValIleSerValIleThrGlySerValPhe 247
 Db 65 TTTCACCTTCAGTTACATATCTATTTATGGTAAATCTGTAAATACTGGGTTCAGTTT 124
 Qy 248 TrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrVal 267
 Db 125 TGGCAATGTCTGCTCTTATCTATCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 184
 Qy 268 PheIleLeuLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSer 287
 Db 195 TTTATCATCAATCTTATTCACCTGATATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 244
 Qy 288 LysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGln 307
 Db 245 AAAAGAGCTACATAGCATATAGCATTTTCTACATTTGTGGTTTATATATATATATATG 304
 Qy 308 IleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyVal 327
 Db 305 ATACCTTTTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGACCTTGCAGGTGC 364
 Qy 328 PheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrIleuArgAspArgLeuThrIysGln 347
 Db 365 TTTCATTTGCTGCAAGCTTAANCCTTTCTTCAGTATCTGAGAC-CGATTACCACCAAC 423

Qy 348 GluPheClnThrLeuPhePheLeuGlyValSerLeu-AlaAlaGlyAlaValPheLeuSe 367
 Db 424 GAGTTCAGACCCCTTTTCNTTTTGGGGGATACACTACTCAGNGCTGGGTCTCTGANGCAT 483
 Qy 367 rValIleTyrLeuThrTyrThrGlyTyrIleAlaPro---Trp 380
 Db 484 TGNATC-----NGGTACATTCCTCCCTCGATGG 511

RESULT 12

US-09-270-767-1318/c
 ; Sequence 1318, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Honburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1318
 ; LENGTH: 560
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-1318

Alignment Scores:

Pred. No.: 3,68e-56 Length: 560
 Score: 595.50 Matches: 119
 Percent Similarity: 78.38% Conservative: 26
 Best Local Similarity: 64.32% Mismatches: 33
 Query Match: 13.53% Indels: 7
 DB: 4 Gaps: 5

US-10-028-384-2 (1-826) x US-09-270-767-1318 (1-560)

Qy 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHis 419
 Db 558 ATTGCATCCGTGTCGGAGCATCAGCCACCACTTGTTCTCTCTCTCTCTCTCTCTCTCT 499
 Qy 420 IleLeuValCysThrPheProAlaGlyLeuTyrPheCysIleLeuAsnIleAsnAspGlu 439
 Db 498 ATCATGTTGGCCCTTCCCACTGGGAGTGTGTCTACTGCATCAGCAGATCAAGCAGAG 439
 Qy 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 Db 438 CGCGTTTCGTGGTGTGTACGCCATCAGTCGGTTTACTTCGCTGGTGTGTGTGTGTGTGT 379
 Qy 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 Db 378 TTGATGTTGACCTCAGCCCGTGTGTGTCATGCTGGCGGGAGTGGCTTTTTCGGGACTG 319
 Qy 480 PheGluHisTyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSer 498
 Db 318 TTGGATGTGTCTCTGCAAGAGGATTCGTCTAAGCAATGGGCACAGCCATAAGCGCAGCC 259
 Qy 499 SerAspGluAsp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGly 515
 Db 258 ACCGAAGTGGATGAAGCTGAGGATTCCTCATGTAAGAAGAAGACGCTGTACGACAGCTGC 199
 Qy 516 LysValArgLysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsn 534
 Db 198 AAGCTG---AAGCATGCTACTAAGCATGATGCCAGCAGGATCTGGCGTTCAGCTCCAC 142
 Qy 535 IleLysSerIleValThrMetLeuMetLeuMetMetMetPheAlaValHisCys 554
 Db 141 CTGAAGAGATTTGTTATTTTGGCCGTTCTTAATGCTGTTGATGATGTTGCTGCTGCAC 82
 Qy 555 ThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn--- 573
 Db 81 ACGTGGGTGACCAAGCAATGCCTACTCCAGTCCCTCCATTGCTTGGCTTCCACACACAGT 22

US-10-028-384-2 (1-826) X US-09-270-767-15440 (1-1660)

| | | | | | | | | | | | | | | | | | | | | | | |
|------|----|-------|------|-------|------|------|------|-------|------|------|------|------|------|-------|-----|-----|-----|-----|-----|-----|------|-----|
| 553 | QY | HIS | CYS | Thr | Trp | Val | Thr | Ser | Asn | Ala | Tyr | Ser | Ser | ----- | Pro | Ser | Val | Val | Leu | --- | 569 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 713 | DB | CAT | TGC | ----- | CTT | GCA | ACG | AGC | GTAT | GTG | CACT | ATT | TGT | TTC | CA | AGT | TAT | TAT | TAT | CA | 766 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 570 | QY | ----- | Ala | Ser | Tyr | Asn | His | Asp | Gly | Thr | Arg | Asn | Leu | Leu | Asp | Asp | Phe | Arg | Glu | Ala | 587 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 767 | DB | TAG | CGT | CTT | CTG | TTC | CCAT | ----- | TCC | CGA | CA | ATT | T | TAG | CA | GAT | TTC | CAG | AGG | CGT | 820 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 588 | QY | Tyr | Phe | Trp | Leu | Arg | Gln | Asn | Thr | Asp | Glu | His | Ala | Leu | Val | Met | Ser | Trp | Trp | Asp | Tyr | 607 |
| | | | | | | | | | | | | | | | | | | | | | | |
| 821 | DB | TAC | TAC | TGG | CTT | TTC | GAGA | CA | CTG | CCG | ATG | ATG | CTC | CGG | TAT | TAT | CTC | TGT | TGG | AT | TAC | 880 |
| | | | | | | | | | | | | | | | | | | | | | | |
| 608 | QY | Gly | Tyr | Gln | Leu | Ala | Gly | Met | Ala | Asn | Arg | Thr | Thr | Leu | Val | Asp | Asn | Thr | Trp | Asn | 627 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 881 | DB | GGAT | ACCA | CAG | ATAG | CGG | GAAT | TGG | CAAA | CAGA | CA | CGCT | AGT | AGT | CGA | TAT | ATA | AT | AT | CGT | 940 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 628 | QY | Asn | Ser | His | Leu | Ala | Leu | Val | Gly | Leu | Ala | Met | Ser | Ser | Asn | Glu | Thr | Ala | Ala | Tyr | Leu | 647 |
| | | | | | | | | | | | | | | | | | | | | | | |
| 941 | DB | ATA | TGT | CA | TAG | CGCT | TGGT | TGG | CA | AGG | CA | TGT | CTT | CA | CCG | AGA | GAG | ACT | CT | AG | 1000 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 648 | QY | Ile | Met | Arg | Thr | Leu | Asp | Val | Asp | Tyr | Val | Leu | Val | Ile | Phe | Gly | Gly | Val | Ile | Gly | Tyr | 667 |
| | | | | | | | | | | | | | | | | | | | | | | |
| 1001 | DB | ATT | T | AG | CA | TCT | CTT | GAC | TGG | CA | TAC | TAC | T | T | T | T | T | T | T | T | 1060 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 668 | QY | Ser | Gly | Asp | Asp | Ile | Asn | Leu | Ser | Phe | Leu | Trp | Met | Val | Ile | Ala | Gly | Gly | His | Pro | 687 | |
| | | | | | | | | | | | | | | | | | | | | | | |
| 1061 | DB | TCT | GCG | ATG | ATAT | CA | CA | AGT | TCC | GTG | ATG | TCC | GAAT | T | GCT | GAG | GAG | AGC | AT | CCC | 1120 | |
| | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6

```

US-09-513-999C-1965
/ Sequence 1965, Application US/09513999C
/ Patent No. 6783361
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Ducluet, A.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: Expressed Sequence
/ Patent No. 6783361
/ FILE REFERENCE: 59.U2.REG
/ CURRENT APPLICATION NUMBER: US/09/513.9
/ CURRENT FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/122,487
/ PRIOR FILING DATE: 1999-02-26

```

; NUMBER OF SEQ ID NOS: 36681

```

; SOFTWARE: Patent.pm
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
; US-03-513-999C-1965

```

| Alignment Scores: | |
|------------------------|----------|
| Pred. No.: | 1.08e-90 |
| Score: | 907.00 |
| Percent similarity: | 100.00% |
| Best Local Similarity: | 100.00% |
| Query Match: | 20.61% |
| DB: | 4 |
| Length: | 507 |
| Matches: | 168 |
| Conservative: | 0 |
| Mismatches: | 0 |
| Indels: | 0 |
| Gaps: | 0 |

US-10-028-384-2 (1-826) X US-09-513-999C-1965 (1-507)

| | | | |
|-----|----|--|-----|
| 562 | QY | TyrSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnLeu | 581 |
| | | | |
| | | | |
| 2 | DB | TACTCTAGTCCAAGTGTAGTCCCGGCTCATCAATCATGATGGCACCAGGAATATCTTA | 61 |
| | | | |
| | | | |
| 582 | QY | AspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgVal | 601 |
| | | | |
| 62 | DB | GATGATTTTACAGAAAGCTTACTTTTGGCTAAGCCAAATATACAGATGAACATGCACGAGTA | 121 |
| | | | |
| | | | |
| 602 | QY | MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuVal | 621 |
| | | | |
| | | | |
| 132 | DB | ATGTCCTGGTGGGATTATGGCTATCAGATAGCTGGGAATGGCTAAATAGAATACTACGTTGGTG | 181 |
| | | | |
| | | | |
| 622 | QY | AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIleValAlaMetSerSerAsn | 641 |
| | | | |
| 182 | DB | GATTAATACACCTCGGAATACACCCCATAGCACTGGTGGAAAGACTATGCTTCTTAAT | 241 |
| | | | |
| | | | |
| 642 | QY | GluThrAlaAlaTyrIleIleMetArgThrLeuAspValAspTyrValLeuValIlePhe | 661 |
| | | | |
| 242 | DB | GAACACAGCAGCCTATAAAATCATGAGAGACTCTAGATGTAGATTATGTTTGGTGTATTTTT | 301 |
| | | | |
| | | | |
| 662 | QY | GlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIle | 681 |
| | | | |
| 302 | DB | GGAGGGGTATTGGCTATTCTGGTGATGATATACAAATTTCTCTGGATGGTTTAGATA | 361 |
| | | | |
| | | | |
| 682 | QY | AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGlu | 701 |
| | | | |
| 362 | DB | GCTGAAGGAGAAACATCCCAAGACATTCGGAAAGTGACTATTTTACCACCACAGGAGAA | 421 |
| | | | |
| | | | |
| 702 | QY | PheArgValaAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSer | 721 |
| | | | |
| 422 | DB | TTCCGTGTAGCAACAGCAGATCCCTACTTTGTGATTGCTTATGATATATAATGTCA | 481 |
| | | | |
| | | | |
| 722 | QY | TyrTrpArgPheGlyGluMetGln | 729 |
| | | | |
| 482 | DB | TACTACAGATTGGAGAAATGCCAG | 505 |

RESULT 7

RESOLLI /
US-09-248-796A-3089
Sequence 3089, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208

Qy 202 AlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAsp 221
 Db 861 GCGCTTTGATATCCATCGTTCGGGGTATATCTCCGATCCGTCGGGATCGTACGAC 920
 Qy 222 AsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrLeuTrpValIleSerVal 241
 Db 921 AATGAAGCAATCCCATTTCTGCATGCTCTTCCACTACTATTGTGGATCAAGCGGTA 980
 Qy 242 LysThrGlySerValPheThrMetCysCysLeuSerTyrPheTyrMetValSer 261
 Db 981 AAGACGGGACGATCTTTGGTGGGTATGTCGGCATTTGGCGCTACTTCTATATGTCCTCC 1040
 Qy 262 AlaTrpGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeu 281
 Db 1041 TCGTGGGTGGCTATGCTTCTCCGATTTACCTTAATCCCGCTGCAGCTGCTGGCGGTATG 1100
 Qy 282 LeuMetGlnArgTyrSerIleValIleAlaTyrSerThrPheTyrIleValGly 301
 Db 1101 ATCACCGGAGCTTTCTCGCAGGATCTACATAGCATACAGCAGCCTACTACTGCTCGGC 1160
 Qy 302 LeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHis 321
 Db 1161 ACCATTCTGTCGATGAGATCTGTTGTGGGATTCACACCCATCCAGAGCTCCGAACAC 1220
 Qy 322 MetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArg 341
 Db 1221 ATGCTGGCACTGGGAACCTTTGGCCTGTGCCAGATTACAGCTTTCTGTCGACTATCTGGC 1280
 Qy 342 AspArgLeuThrIleGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 361
 Db 1281 TCGGCGATCCCAAGGATCATCTCGATCGTCTCTTCAAGAGCTTGGTTCCAGTCTTTTG 1340
 Qy 362 GlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGly 375
 Db 1341 ACTGTGTGTTCTGTCGGGTACCTGCTCAGCTTACCGGG 1382

RESULT 4

US-09-270-767-158
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 158
 ; TYPE: DNA
 ; LENGTH: 1660
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Alignment Scores:
 Pred. No.: 2,03e-97 Length: 1660
 Score: 976.00 Matches: 188
 Percent Similarity: 79.29% Conservative: 34
 Best Local Similarity: 67.14% Mismatches: 44
 Query Match: 22.18% Indels: 14
 Gaps: 6

US-10-028-384-2 (1-826) x US-09-270-767-158 (1-1660)

Qy 553 HisCysThrTrpValThrSerAsnAlaTyrSerSer-----ProSerValValLeu--- 569
 Db 713 CATTTGC-----CTTGGCAACACCGCTATGTGACTATTGTTCCAGCTATTATTCAC 766
 Qy 570 -----AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAla 587
 Db 767 TAGGTCCTTCTGTTTCCAT-----TCCGGCAACATTTTAGACCATTTTCAGAGGCT 820
 Qy 588 TyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyr 607

Db 821 TACTACTGCTTTCCAGAACACCTGCCGATGATGCTCGGTATGTCTTGGTGGATTAC 880
 Qy 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsn 627
 Db 881 CGATACCATAGCGGGAATGCAACAGACGAGCTAGTGGATAATAATACGTGGAAC 940
 Qy 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
 Db 941 AATAGTCACATAGCTGCTGGCAAGCAATGTCTTCAACCGGAGAGAACTCTCAGAA 1000
 Qy 648 IleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyr 667
 Db 1001 ATTATGACATCTCTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGTAT 1060
 Qy 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisPro 687
 Db 1061 TCTGGCGATGATCAACAAAGTTCTCTGGATGGTCCGAATGCTGAGGAGACATCCC 1120
 Qy 688 LysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAla 707
 Db 1121 AAGGACATTAAAGAAAGCGATTACTTACCGACCGCGGTGAATTCAGGGTAGATGCCAA 1180
 Qy 708 GlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGlu 727
 Db 1181 GGTGCTCCCGCCCTGCTCAACTGCCTTATGTACAAATTAAGCTACTACAGATTCCGGAA 1240
 Qy 728 MetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGly 747
 Db 1241 TTGAAGTTGGACTACAGAGGTCATCTGGATATGATGCACACGTAACCGCTCATTTGG 1300
 Qy 748 AsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuVal 767
 Db 1301 AATAAGGACTTCGATCTGACTTACCTCGAGGAGGCTACACACAGAACTCGGCTTGT 1360
 Qy 768 ArgIleTyrLysValIleAlaProAsp-----AsnArgGluThrLeuAspHisLysPro 785
 Db 1361 CGCATCTATAGGTGAAGAGCGCATGATGATTCATATAGCATCATCAAGACCAAGGAG 1420
 Qy 786 ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys 805
 Db 1421 AGAAGC-----ATTCCTCCAGAAACTTCATTTGAGAAAGAACTCTAAGCGTCCG 1471
 Qy 806 ArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThr 825
 Db 1472 AAGGCTACATACGAAACCGCGGTTGTTTAAGGGAAACGAACTTGAATAAACC 1531

RESULT 5

US-09-270-767-15440
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

Alignment Scores:
 Pred. No.: 2,03e-97 Length: 1660
 Score: 976.00 Matches: 188
 Percent Similarity: 79.29% Conservative: 34
 Best Local Similarity: 67.14% Mismatches: 44
 Query Match: 22.18% Indels: 14
 Gaps: 6

Thu Dec 16 16:25:06 2004

Db 2110 AAACCAAGAGGTCATTAAGAGAGACCTCAATTA 2142

RESULT 2

US-09-270-767-12331

; Sequence 12331, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 12331:

; LENGTH: 900

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-12331

Alignment Scores:

Pred. No.: 5,09e-118 Length: 900
Score: 1158.00 Matches: 213
Percent Similarity: 90.5% Conservative: 26
Best Local Similarity: 80.6% Mismatches: 25
Query Match: 26.31% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-12331 (1-900)

```

QY 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
Db 109 AGCAAGGTGGTGGTATACAGAGCTATACCTTCGCCATCTCTTAATCGCTGGCTG 168
QY 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
Db 169 GCGGATTTCTCTCGCTTCGCGGTATCGCTTCGAGTTCGATTCATTCATGAGTTT 228
QY 103 AppProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTyrluPhe 122
Db 229 GATCCGTGTTCACTACCGGCCACCGCTACATCGTGCAGATGGTGGTACAACTTC 288
QY 123 LeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGlyThrVal 142
Db 289 CTCACTGTTTCAGCAGCGCGCATGTTATCGCTCGGAGATTTGTGGCGGTACCGTC 348
QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrIleAsnIle 162
Db 349 TATCCGGGCTGATGATTACGTCCGCGGGAATCCATTGGCTCTGCACGTACTCAACATA 408
QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
Db 409 CCGGTCCATATTCGTGACATCTCGGTTCCTGGCGCCGATCTTTCAGTGGCTGACCTCC 468
QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuLeuAla 202
Db 469 ATCTCCACCTACCTGTCACCAAGAGCTGTGGTCCGCGGCGCGCTCTTCGCCGCC 528
QY 203 CysPheIleAlaIleValProGlyTyrlleSerArgSerValAlaGlySerPheAspAsn 222
Db 529 AGCTTCATGCCATCGTGGCTGCTACATCAGTAGTCTGGTGGTGGTGGTGGTGGTGGT 588
QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrlleTrpValbySerValIys 242
Db 589 GAGGCAATGCCATATTCGCTTCGAGTTCACCTACTTCTGTGGTGGTGGTGGTGGTGGT 648
QY 243 ThrGlySerValPheTrpMetCysCysLeuSerTyrllePheTrpMetValSerAla 262
Db 649 ACTGGATCCGTGTTCTGGTCCGCGCGCGCGCTTGTCTTACTTCTACATGGTGGTCC 708
QY 263 TrpGlyGlyTyrlleValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
Db 709 TGGGTGGCTACGTGTTCACTACCACTGATACCTGATACCTGACGCTCTTCTGCTGCTCAT 768

```

```

QY 283 MetGlnArgTyrlleSerLysArgValTyrlleAlaTyrlleSerThrPheTyrlleValGlyLeu 302
Db 769 ATGGCAGGTACTCGCGCGCTCTGCTGACCACTACAGCACTTCTATCATCTGGGACTG 828
QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
Db 829 CTGTTCTTCATCCAGATCCCTTCGTGGGATTCACCGATACGACCACTGACACATG 888
QY 323 AlaAlaAlaGly 326
Db 889 GCTGCGCTGGGA 900

```

RESULT 3

US-09-270-767-11648

; Sequence 11648, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 11648

; LENGTH: 1386

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-11648

Alignment Scores:
Pred. No.: 8.32e-104 Length: 1386
Score: 1032.50 Matches: 196
Percent Similarity: 74.55% Conservative: 53
Best Local Similarity: 58.68% Mismatches: 84
Query Match: 23.46% Indels: 1
DB: 4 Gaps: 1

US-10-028-384-2 (1-826) x US-09-270-767-11648 (1-1386)

```

QY 43 LysAlaAlaGlyGlyAlaAlaProProLysProAlaPro---AlaGlyLeuSerGlyGly 61
Db 381 AAGCCCGCGAGCGGAGAGAGAGCAAAATGACGCTGAGCGTGGGGCGGCTGCCAGGGA 440
QY 62 LeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrp 81
Db 441 CTGCTCACCTGGGACAAACAGGAGACCTGGTCAAGCTGGCCATCTCATCTCGCAGCG 500
QY 82 LeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHisGlu 101
Db 501 GTTATATCAATTCACACGCTTGTCTCTGTGCTCGGATTCGAAAGCGTATCCATGAG 560
QY 102 PheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSerHisGlyPheTyrlu 121
Db 561 TTGATCCGCTACTCACTACCCACCAACGCGGCTTCGCGGAGCAGGCGCTTTTACAAG 620
QY 122 PheLeuAsnTrpPheAspGluArgAlaTrpTyrlleProLeuGlyArgIleValGlyThr 141
Db 621 TTCACAACTGGTTCGATGACCGCGCTGGTATCCCTGGCGCGCATCATCGCGCGCAC 680
QY 142 ValTyrlleProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsn 161
Db 681 ATCTATCCCGGCTGATGCTACCTCGGCGGCGCTGTACCGCTGATGTGGCTGCTCAAT 740
QY 162 IleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThr 181
Db 741 GTGACCATCGACATACGAAACGCTGGCTCTCTCGCGCGCTTCTCTCTCGCTGACC 800
QY 182 SerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 201
Db 801 ACGTGTGACCTTACGCCCTCACAAAGAGATACACAGCACTGGAGCTGGACTGGTGGCC 860

```

Qy 88 ArgLeuPheAlaValIleAlaArgPheGluSerIleIleHisGluPheAspProTrpPheAsn 107
Db 94 CGTTGTTGTCAGTCATCAAAATTTGAGTCTATTATCCATCAATGAACCCCTGGTTCAAT 153
Qy 108 TyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAsp 127
Db 154 TATAGGCTACCAATATCTCGTCAACAATTCGTTTACAAAGTTTGAACCTGGTTGAC 213
Qy 128 GluArgAlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMet 147
Db 214 GACGTCACCTGGTACCCCTCGAAGGGTTACTGGAGGACTTTATATCCTGGTTGARG 273
Qy 148 IleThrAlaGlyLeuIle---HisTrpIleLeuAsnThrLeuAsnIleThrValHisIle 166
Db 274 ACAGTATGCTTCATCTGCAGCCCTGCGCAACTGGTGGCTTGGCTTCCATTCACATC 333
Qy 167 ArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPhe 186
Db 334 AGAAACGTTTGTGTGCTATTTCGGCCACTATTTCCTGGGGTCAACCGCTGGCGACTTAC 393
Qy 187 LeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuAlaAlaCysPheIleAla 206
Db 394 GAATTTACGAAAGAGATTAAAGATGCCAGCGCTGGCTTTGGCTGCTGCTTTATAGCC 453
Qy 207 IleValProGlyTyrIleSerArgSerValAlaGlySerPheAsnGlnGlyIleAla 226
Db 454 ATTCCTCCCGGTTATATATCTAGATCAGTGGCGGGTCTCTAGATTAATGAGGCCATTGCC 513
Qy 227 IlePheAlaLeuGlnPheThrTyrTyrLeuTrpValTyrSerValIysThrGlySerVal 246
Db 514 ATTACACTATTATGCTTTCATCTTTCGATTAAAGCCCAAGAGCTGGCTCTATC 573
Qy 247 PheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyr 266
Db 574 ATGCAGCAAGCTGTGAGCTTATTCTACTTCTCATGCTGTGCTGCTTGTGCTGCTGCTATC 633
Qy 267 ValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyr 286
Db 634 GTGTTTCATCACCACCTTGTATCCCATCTCATGTCTTTCGCTGATTGTTGAGGCGAGATAT 693
Qy 287 SerIysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuSerMet 306
Db 694 TCGTCCAAACGTATTCTGCTACCCACTTGTGCTGCTTGTGCTGCTTGTGCTGCTGCTATC 753
Qy 307 GlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGly 326
Db 754 CAGATCCCATTTGTCGTTTCTTCTACCTATCAGTCTTAACGACCATGCGCGCTTGGCT 813
Qy 327 ValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIys 346
Db 814 GTTTTCGGTTTGTATTCAGATTGTCGCTTTCGCTGCTTTCGCTGCTTTCGCTGCTGCTGCT 873
Qy 347 GlnGluPheGlnThrLeuPhePhe-----LeuGlyValSerLeuAlaAlaGlyAlaVal 364
Db 874 GCTAAGTTTAAAGTCATCATGATGTTTCTGTTTTCGCTTTCGCTTTCGCTTTCGCTGCTGCT 933
Qy 365 PheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPhe 384
Db 934 GGACTTTTCGCC-----TTGACCTATATGGGGTGTGATTCGCCCTTGGACTGCTGCTGCTGCT 987
Qy 385 TyrSerLeuTrpAspThrGlyTyrAlaValIleHisIleProIleIleAlaSerValSer 404
Db 988 TATTCGTTATGGGATACCAACTACGCAAGATCCACATTCCTATCATTCGCTCCGTTTCC 1047
Qy 405 GluHisGlnProThrTrpValSerPhePheAspLeuHisIleLeuValCysThr 424
Db 1048 GAACATCAACCCGTTTGTGGCCCGCTTCTTCTTGTATACCCACTTTTGTATCTGGCTA 1107
Qy 425 PheProAlaGlyLeuTrpPheCysIleIleAsnIleAsnAspGluArgValPheValAla 444
Db 1108 TTCCCGCGCGGTATTCTTCTACTATTCTCGACTTGAAGACGAGCAGCTTTTGTGCTATC 1167
Qy 445 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 464

Db 1168 GCTTACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1227
Qy 465 ThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeu 484
Db 1228 ACACCACTCATCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1287
Qy 485 GlyAspAspMetIysArgGluAsnProProValGluAspSerSerAspGluAspAspLys 504
Db 1288 -----GATTTCAG-----ACAACTGAC----- 1305
Qy 505 ArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGln 524
Db 1306 -----CGCAATACGCCATCAACCT 1326
Qy 525 GluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeu 544
Db 1327 CGCGCACTA-----CTGGCCAAATTTGATTTTCCGGATCATTCATC 1368
Qy 545 MetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSer 564
Db 1369 TTTTATTTCTATCTTTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1428
Qy 565 ProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPhe 584
Db 1429 CTTCTGTTGTTTGGCATCAAAACCCAGATGTAATTTGGCTTGTATCGACGACTTC 1488
Qy 585 ArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrp 604
Db 1489 AGGGAAGCTGCTATTGTTAAGAATGAATCTGATGAGGACAGTAAAGTTGCGAGCTGG 1548
Qy 605 TrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsn 624
Db 1549 TGGATTAACGTTTACCNAATTTGGTGCATGCGACAGACACCACTTTAGTCGATACAC 1608
Qy 625 ThrTrpAsnAspSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAla 644
Db 1609 ACGTGGAACAATACTCACATCGCATCGCTTGTGTAAGCCATGCTGCTCCCTGAAGAGAA 1668
Qy 645 AlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyVal 664
Db 1669 TCTTACGAATTTTAAAGAGCATGATGCTGATTTGCTTGGTTCATCTTTGGTGTCTA 1728
Qy 665 IleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGly 684
Db 1729 ATTGGGTTTGGTGTGATGACATCAACAATTTCTTGTGATGATCAGATTTAGCGAGGA 1788
Qy 685 GluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgVal 704
Db 1789 ATCTGCGCAGAGAGATTAAGAGCGTGTATTTCTATACGACAGAGAGAGATACAGACTA 1848
Qy 705 AspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArg 724
Db 1849 GATGCAAGGGCTTCTGAGACCATGAGAACTCGCTACTTTTACAGATGCTCTACAAAGAT 1908
Qy 725 PheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAla 744
Db 1909 TTCCCACTTA-----TTCAATGTTGCCAAGCCACTGACAGAGTGGTCAACA 1959
Qy 745 GluIleGlyAsnLysAspIle-----LysPheLysHisLeuGluGluAlaPheThrSerGlu 763
Db 1960 ATGATCACACCATTTAGACGCTCCACCATCTAGACTACTTTCGAGAAAGTTTACTTCCGAA 2019
Qy 764 HistTrpLeuValArgIleTyrLysValLysAlaProAspAsnArg-----GluThrLeuAsp 782
Db 2020 AACTGATGTTAGATATATCATTTGAAGAGGATGATGCCCAAGGTAGACACTTTGAGG 2079
Qy 783 HisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysThrThr 802
Db 2080 GACGTTGCTGAGTTAACC-----AGTCTTCTTACG 2109
Qy 803 LysArgLysArgGlyTyrIleLysAsnLysLeu 813

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - nucleic search, using frame_plus_p2n model
Run on: December 14, 2004, 18:57:05 ; Search time 160.664 Seconds
(without alignments)
3654.277 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 4401
Sequence: 1 MAFPSAPESKSHKSLNNSPW.....GYKNKLVFKGKISKKTV 826

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0
Searched: 824507 seqs, 355394441 residues
Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPFO.spoel/US10028384/runat_14122004_131520_8052/app_query.fasta_1.3740
-DB=issued Patents NA -QFMT=fastap -SUFFIX=arni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.1 259 @runat_14122004_131520_8052 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUTS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|---------------------|
| 1 | 2021.5 | 45.9 | 2157 | 4 | US-09-614-221A-318 |
| 2 | 1158 | 26.3 | 900 | 4 | US-09-270-767-12331 |
| 3 | 1032.5 | 23.5 | 1386 | 4 | US-09-270-767-11648 |
| 4 | 976 | 22.2 | 1660 | 4 | US-09-270-767-158 |
| 5 | 976 | 22.2 | 1660 | 4 | US-09-270-767-15440 |
| 6 | 907 | 20.6 | 507 | 4 | US-09-513-999C-1965 |
| 7 | 831 | 18.9 | 867 | 4 | US-09-248-796A-3089 |
| 8 | 806.5 | 18.3 | 1660 | 4 | US-09-270-767-158 |
| 9 | 806.5 | 18.3 | 1660 | 4 | US-09-270-767-15440 |
| 10 | 742.5 | 16.9 | 914 | 4 | US-09-270-767-12956 |
| 11 | 704 | 16.0 | 616 | 3 | US-09-328-111-332 |
| 12 | 595.5 | 13.5 | 560 | 4 | US-09-270-767-1318 |

| | | | | | | | |
|---|----|-------|------|-----|---|----------------------|-------------------|
| C | 13 | 595.5 | 13.5 | 560 | 4 | US-09-270-767-16500 | Sequence 16500, A |
| | 14 | 553 | 12.6 | 503 | 4 | US-09-270-767-27262 | Sequence 27262, A |
| | 15 | 544 | 12.4 | 594 | 4 | US-09-248-796A-3090 | Sequence 3090, Ap |
| | 16 | 469 | 10.7 | 487 | 3 | US-09-385-982-213 | Sequence 213, App |
| | 17 | 427 | 9.7 | 245 | 4 | US-09-513-999C-21090 | Sequence 21090, A |
| | 18 | 410 | 9.3 | 250 | 4 | US-09-513-999C-1438 | Sequence 1438, Ap |
| | 19 | 406 | 9.2 | 487 | 4 | US-09-702-705-1655 | Sequence 1655, Ap |
| | 20 | 406 | 9.2 | 487 | 4 | US-09-736-457-1655 | Sequence 1655, Ap |
| | 21 | 406 | 9.2 | 487 | 4 | US-09-614-124B-1655 | Sequence 1655, Ap |
| | 22 | 406 | 9.2 | 487 | 4 | US-09-671-325-1655 | Sequence 1655, Ap |
| | 23 | 406 | 9.2 | 487 | 4 | US-09-658-824-1655 | Sequence 1655, Ap |
| | 24 | 398 | 9.0 | 307 | 4 | US-09-702-705-375 | Sequence 375, App |
| | 25 | 398 | 9.0 | 307 | 4 | US-09-702-705-1271 | Sequence 1271, Ap |
| | 26 | 398 | 9.0 | 307 | 4 | US-09-736-457-375 | Sequence 375, App |
| | 27 | 398 | 9.0 | 307 | 4 | US-09-736-457-1271 | Sequence 1271, Ap |
| | 28 | 398 | 9.0 | 307 | 4 | US-09-614-124B-375 | Sequence 375, App |
| | 29 | 398 | 9.0 | 307 | 4 | US-09-614-124B-1271 | Sequence 1271, Ap |
| | 30 | 398 | 9.0 | 307 | 4 | US-09-671-325-375 | Sequence 375, App |
| | 31 | 398 | 9.0 | 307 | 4 | US-09-671-325-1271 | Sequence 1271, Ap |
| | 32 | 398 | 9.0 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| | 33 | 398 | 9.0 | 307 | 4 | US-09-658-824-375 | Sequence 375, App |
| | 34 | 398 | 9.0 | 307 | 4 | US-09-658-824-1271 | Sequence 1271, Ap |
| | 35 | 392 | 8.9 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| | 36 | 392 | 8.9 | 307 | 4 | US-09-736-457-588 | Sequence 588, App |
| | 37 | 392 | 8.9 | 307 | 4 | US-09-614-124B-588 | Sequence 588, App |
| | 38 | 392 | 8.9 | 307 | 4 | US-09-671-325-588 | Sequence 588, App |
| | 39 | 392 | 8.9 | 307 | 4 | US-09-589-184-588 | Sequence 588, App |
| | 40 | 392 | 8.9 | 307 | 4 | US-09-658-824-588 | Sequence 588, App |
| | 41 | 359 | 8.2 | 292 | 4 | US-09-313-294A-4442 | Sequence 4442, Ap |
| | 42 | 332 | 7.5 | 291 | 4 | US-09-313-294A-4834 | Sequence 4834, Ap |
| | 43 | 326 | 7.4 | 414 | 4 | US-09-513-999C-271 | Sequence 271, App |
| | 44 | 297 | 6.7 | 302 | 4 | US-09-313-294A-6869 | Sequence 6869, Ap |
| | 45 | 295 | 6.7 | 269 | 4 | US-09-313-294A-906 | Sequence 906, App |

ALIGNMENTS

RESULT 1
US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunananda, Balasulojini
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae

Alignment Scores:
Pred. No.: 7,21e-213 Length: 2157
Score: 2021.50 Matches: 397
Percent Similarity: 68.5% Conservative: 118
Best Local Similarity: 52.8% Mismatches: 183
Query Match: 45.9% Indels: 53
DB: 4 Gaps: 11

US-10-028-384-2 (1-826) x US-09-614-221A-318 (1-2157)

Qy 68 TTPGInSerLeuLeuSerPheThrIleuPheLeuAlaGlyPheSerSer 87
Db 34 TTTCAGACCACCTCAAGCTGTCATCTTCGTGGCGATTTCGGGTGCATATCATCA 93

| | | | |
|----|------|---|------|
| QY | 193 | TrpAenGlnGlyAlaGlyLeuLeuAlaAalaCysPheileAlaileValProGlyTyrile | 212 |
| Db | 439 | AAGGATTCTAGTCAGGAGTATTGGCAGCTATATTATGGGAGTTGCCCCAGGTTATATT | 498 |
| QY | 213 | SerArgSerValAlaGlySerPheAspAenGluGlyileAlailePheAlaLeuGlnPhe | 232 |
| Db | 499 | TCAGATCAGTGGCTGGTCTTATGATATATGAACATTTGCCATTACTTTATTAATGGCA | 558 |
| QY | 233 | ThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCysCys | 252 |
| Db | 559 | ACATTTTATTCTCGATTAATCAATCAAAATGGGTTTCAGTTTCTATGCCACATTGACA | 618 |
| QY | 253 | CysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheileleAsnLeu | 272 |
| Db | 619 | GCATTATTCTATATAGTTAGTCTTGGGTGGATATGTTTCATTACCAATTG | 678 |
| QY | 273 | IleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrile | 292 |
| Db | 679 | ATTCCATTACACGATTGTGCTTGAATTTATGGGTCGTATAATGCCAACTTTTACACT | 738 |
| QY | 293 | AlaTyrSerThrPheTyrileValGlyLeuileLeuSerMetGlnileProPheValGly | 312 |
| Db | 739 | GCTTATACATCAGTATGCTTGGGTACTTTGGCATCAATGCAGATTCCATTGTTGGG | 798 |
| QY | 313 | PheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGln | 332 |
| Db | 799 | TTTTTACCATTAGATCAATGATCATATGGCTGCATTAGGAGTATTTGGATTGTTACAA | 858 |
| QY | 333 | AlaTyrAlaPheLeuGlnTyrLeuA:GAspArgLeuThrLysGlnGluPheGlnThrLeu | 352 |
| Db | 859 | TTAGTGGCTTTTGTGATTTATGTTAAATCAAAAGTTCCAACCAACAAATTTAAATCATTC | 918 |
| QY | 353 | PhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValileTyrLeuThr | 372 |
| Db | 919 | TTGATAGTTCCATTGCTACTTGTGTGGATTAGGTATGGTGGATTTATGGATTACCA | 978 |
| QY | 373 | TyrThrGlyTyrileAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyr | 392 |
| Db | 979 | GCAATGGGTGGATTGCTCTCGACAGTAGATTTTATCTTATGGGATACAAATAT | 1038 |
| QY | 393 | AlaLysIleHisIleProIleleAlaSerValSerGluHisGlnProThrThrTrpVal | 412 |
| Db | 1039 | GCCAAAGATTCAATCCAAATTATGCTTCTGTGTTCTGAACATCAACCTACTGCTGGCCA | 1098 |
| QY | 413 | SerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCys | 432 |
| Db | 1099 | GCATTCCTTTTCGAACTAGTAGCTATTTGGTTATCCCGCTGGTATCTATTATGT | 1158 |
| QY | 433 | IleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaileSerAlaValTyr | 452 |
| Db | 1159 | TTCCAAGAAATGAAGGATGAACAGTTCATTATCATTTACATGATTTGTTGTTCTAT | 1218 |
| QY | 453 | PheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSer | 472 |
| Db | 1219 | TTTGCTGGTGCATGGTAAGATTGATGTGACTTTGACTCCAGTCAATTGTGTGCTGCA | 1278 |
| QY | 473 | AlaIleAlaPheSerAsnValPheGluHisTyrLeuGly- | 486 |
| Db | 1279 | GCAATTGCTTATCTAAATTTGTTGATGCTATTTGGACATGTTGTGATTTTTCACCTGAG | 1338 |
| QY | 487 | AspMetLysArgGluAsnProValGluAspSerSerAspGluAspLysArgAsn | 506 |
| Db | 1339 | AAAGTGGAAAGTATGATGATGACGTTAGTGCAATCCAAAGAGCTCAACCAAAATCA | 1398 |
| QY | 507 | GlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLys | 526 |
| Db | 1399 | AGTTCACAGA---TTTCCAATTGCTGATTTTGCA------ | 1431 |
| QY | 527 | ThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeu | 546 |
| Db | 1432 | -----AAAGTTTGGTTTACTGACATTTACATTTTAC | 1464 |

| | | |
|------|--|------|
| 547 | LeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSer | 566 |
| 1465 | CTTTCTTACTTTTGTTTTCATTTGTACTTGGGTAAACATCGAATGCTTATTCATCACCATCA | 1524 |
| 567 | ValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGlu | 586 |
| 1525 | GTGTGTTTGTAGATCCAGAACCCAGATGGCTCACAACATCATCATGATTATAGAGAA | 1584 |
| 587 | AlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAsp | 606 |
| 1595 | GCCTATTACTGTTTAAAGATGAATACACAGAGAAGATGCCAAAGTTATGGCGCTGGTGGAT | 1644 |
| 607 | TyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrp | 626 |
| 1645 | TATGGTTTATCAAAATCGGGGTATGGCTGTAGAGAACACACACTGTGTGATAACAATACATACG | 1704 |
| 627 | AsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyr | 646 |
| 1705 | ANTACACACATATTGCCACTGTTGGTAAGGCAATGCTCTCCCTGAAGATGTGCGTAT | 1764 |
| 647 | LysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGly | 666 |
| 1765 | GAAAATTTTGAGACAAACACCATGTGATTATGTGTAGTTATATTGTCAGAGGTTATTGGGT | 1824 |
| 667 | TyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHis | 686 |
| 1825 | TATTCTGGTGATGATATTACAAATCTTATGGATGGTGAAGANTGCTGAAGGTATCTCG | 1884 |
| 687 | ProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLys | 706 |
| 1885 | CCTGATCAATCAAGAAAGAGACTACTTTACTGACCAGGAGAAATATAAAGTGGATAAA | 1944 |
| 707 | AlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGly | 726 |
| 1945 | GATGCATCACTGGCAATGAGAATCTTTGTATGTATAAGTTATCGTATCATAGATTCACT | 2004 |
| 727 | GluMetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIle | 746 |
| 2005 | GAATIG-----TTTGAGGTAGATGGTGTGTATAGATTAGAAACCAACAAATC | 2055 |
| 747 | ---GlyAsnLysAspIleLysPheLysHisLeuGluAlaPheThrSerGluHisTrp | 765 |
| 2056 | CCAGCCCAATGAAGTACCGAAATTTGAATGTTGTTGAAGAAGCCTTCACATCAGAAAATGG | 2115 |
| 766 | LeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysPro | 785 |
| 2116 | ATTGTGAGATTTTACAAGTTAAAGATTGGGATAATGTTGGTAGAGATTACATCAAA--- | 2172 |
| 786 | ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys | 805 |
| 2173 | -----CCTACTGCTTTTGAAGATCA-----TCATCCGGCACTTCCAAAAGAAAC | 2217 |
| 806 | ArgGlyTyrIleLysAsnLysLeu | 813 |
| 2218 | AGATCCATAAAGAGACCTAAATTG | 2241 |

Search completed: December 14, 2004, 20:01:33
Job time : 1088.96 secs

QY 266 TyrValPheIleIleAenLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArg 285
 Db 814 TACGTTTTCATCACCACATGATTCATTCGACGCTTGTCTCATTCATTCATGCGCAG 873
 QY 286 TyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSer 305
 Db 874 TTCACACACCGCTTATACCGCTTACTCTCTGCTGTAATGATTCGTAATTCGCTCC 933
 QY 306 MetGlnIleProPheValGlyPheGlnProIleAargThrSerGluHisMetAlaAla 325
 Db 934 ATGACGTCCTCTTGTGGATTCCTCCCATCCGAACCTTCGACACATGCGCGCTTG 993
 QY 326 GlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThr 345
 Db 994 GGTGTTTCGTCCTTGTACGCTCATCGGATTCGTGCAAGTCGTCCGACGACTCGTGCT 1053
 QY 346 LysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPhe 365
 Db 1054 GCGAAGCAATTCACGCTCTCTCAAGCTTTTGTGTCGTCGCTTATTCGCTCATGTTT 1113
 QY 366 LeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProIlePheArgPheTyr 385
 Db 1114 CTGCTCCCTGTCACCTTGTGCTTCTCTGATGATGCGCCCTTCGCTGGAAGATTTAT 1173
 QY 386 SerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGlu 405
 Db 1174 TCTCTTTGGGATCTGCGCTATGCGAGGTCCATGCTCCATATTCCTCCGCTCCGAA 1233
 QY 406 HisGlnProThrThrTrpValSerPhePhePheAspLeuHisIleLeuValCysThrPhe 425
 Db 1234 CACGACCCACCGCTTGGCTCATCTACTTTGACCTCGAATGCTTATCTTCTTTTC 1293
 QY 426 ProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeu 445
 Db 1294 CTTGCGGTGTCCTTGTGTTTCAAGAGCTTTCGCGATGAGCAGATCTTCATCATATT 1353
 QY 446 TyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThr 465
 Db 1354 TATGCGGTTCATGAGCTATTTTGGCGGTGTCATGCTGCTGCTGCTGCTGCTGCTGCT 1413
 QY 466 ProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGly 485
 Db 1414 CTTGTTGTCGTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1470
 QY 486 AspAspMetLysArgGluAsnProProValGluAspSerSerAspGluAspAspLysArg 505
 Db 1471 -----GACCCGCTCATCCCCGAAGCGACGAGGAGCTGCGAG 1509
 QY 506 AsnGlnGlyAsnLeuTyrAspAlaGlyLysValArgLysHisAlaThrGluGlnGlu 525
 Db 1510 TCTCAGCGCAGGTGTTCTCCAAAGTCC---AAGCGGAAGAAGATGCGCGCTGCCACGCC 1566
 QY 526 -----LysThrGluGluGlyLeu--- 531
 Db 1567 AATAAGACGGGTCTCTTTCACAGGTATTTTGGCGGCAAGTCTGCTCCGCGATCTTT 1626
 QY 532 GlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetLeuMetLeuMet 551
 Db 1627 GGTCTCGACATCGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1686
 QY 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSer 571
 Db 1687 CTTCACTGCATATGTCATCTTCAACAGCGTATTTCTTCGCTTTCAGTGTGCTGCTGCTGCT 1746
 QY 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeu 591
 Db 1747 CGAAACCGCGATGTTAGCAAAATATCATTTGATTTCCGAGAGGCTTACTACTGCTGAT 1806
 QY 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
 Db 1807 CGCCAAACACCGCGAGACGAGCGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1866
 QY 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631

Db 1867 GCTGGTATGCTGATCGCCACCCCTTGTGTATAAACAATCTGTAATAACCCACAT 1926
 QY 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
 Db 1927 GCCACAGTTGGTAAGCCATGGCTTCCACGAAGATGCGCATATCCATCTTGGGAAG 1986
 QY 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
 Db 1987 CATGATGTCGATACCTTCTTGTGATCTTGTGGGCTTATTGGGCTACTCTGGTGACAT 2046
 QY 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691
 Db 2047 ATCAACAAGTTTGTGGATGGTATAGGATCTCACAAAGTGAATGCGCTGACGAGTGCGAG 2106
 QY 692 GluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThr 711
 Db 2107 GAAGTCAACACTTACTCAAAAGGGGAGTATGCTGCTGATGACGAGGCGCCACTACT 2166
 QY 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
 Db 2167 ATGAAGAAGTCTCTCATGTACAAATGTCTTACTACCGCTTCCCGAGCTT----- 2217
 QY 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIle 751
 Db 2218 TATGCTGGACACCGGCTCAAGACAGGTTCCAGGCGCAAAATATCCCTCAACAGTGT 2277
 QY 752 LysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLys 771
 Db 2278 ACTCTTGATATCTTCCAGCAAGGTTTACATCCGAAATTTGGATCGTCAGGATCTTACAG 2337
 QY 772 ValLysAlaProAsp-----AsnArgGluThrLeuAspHisLys 784
 Db 2338 GTCAAGAAGAAGATCCCATTTGACAGA-----GACCAACAG 2373
 RESULT 14
 ABT20789
 ID ABT20789 standard; DNA; 2232 BP.
 XX
 AC ABT20789;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Aspergillus fumigatus essential gene #3147.
 XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 OS Aspergillus fumigatus.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287066P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 DR
 XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.

XX 20-FEB-2001; 2001WO-US005171.
 XX 17-FEB-2000; 2000US-0183319P.
 XX 16-MAR-2000; 2000US-0189862P.
 XX 25-MAY-2000; 2000US-0207454P.
 XX 09-JUN-2000; 2000US-0211314P.
 XX 18-JUL-2000; 2000US-0219007P.
 XX 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA Schlegel R, Endege WO, Monahan JE;
 PI WPI; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 PS Claim 1; Page 4632-4633; 11750pp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a compound for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,246-203 Length: 2760
 Score: 2236.00 Matches: 429
 Percent Similarity: 74.20% Conservative: 103
 Best Local Similarity: 59.83% Mismatches: 147
 Query Match: 50.81% Indels: 38
 DB: 5 Gaps: 7
 US-10-028-384-2 (1-826) x ABV24502 (1-2760)
 QY 70 SerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerArgLeu 89
 DB 150 ACACCTTTGAAGCTTCTCATTTCTGTCATGGCTGCTGTATTATCTCTCCACTGTCG 209
 QY 90 -PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrp 109
 DB 210 TTTTGTGCTGCTGAGATTGAAGTTGAAGTTGATTCATGAGTTTGCCTGTTTATATG 269
 QY 109 gSerThrHisHisLeuAlaSerHisGlyPheTyrrGluPheLeuAsnTrpPheAspGlu 129
 DB 270 GACTACCAAGTTTCTGCTGAGGAGGGGTTTATATAATTCATAAAGTGGTTTCATGACG 329
 QY 129 gAlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTyrrProGlyLeuMetIle 149
 DB 330 AGCTTGGTACCTTTGGACGAGATCATATGAGGAGCAATTTACCGAGTTTATGATCAT 389
 QY 149 rAlaGlyLeuIleHisTrpIleLeuAsn-ThrLeuAsnIleThrValHisIleArgAsp 169
 DB 390 CTCGTGCTGAATCTACCATGACTCCATTTTTCACATCCACATCCGATTCGGAATG 449
 QY 169 aICysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeu 189
 DB 450 TCTGTGTGTCTGCGCCCT 509
 QY 189 hArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaValP 209

DB 510 CCAAGAGCTCAAGGATCGAGGGCTGGGCTTCTTCTGCTGCGCATGATGCTGTAGTTC 569
 QY roGlyTyrrIleSerArgSerValIleAlaGlySerPheAspAsnGluGlyIleAlaIlePhe 229
 DB 570 CTGGATATATCTCCGACTCTGTGGCTGCTCTATGATTAATGAAGGATTCGCATCTTTT 639
 QY 229 lAlaLeuGlnPheThrTyrrTyrrLeuTrpValIleSerValIleValIleValIle 249
 DB 630 GCATGCTACTACCTACTACATGATCAAGGAGGAGTAAAGACTGGTTCCATCTGTGGG 689
 QY 249 hrMetCysCysCysLeuSerTyrrPheTyrrMetValSerAlaTrpGlyGlyValPhe 269
 DB 690 CAGCTAAGTGTGCTTCTTCTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 749
 QY 269 lAlaLeuGlnPheThrTyrrIleValGlyLeuIleLeuSerMetGlnIleP 309
 DB 750 TGATCAACTTAATCTCTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 809
 QY 289 rGValTyrrIleAlaTyrrSerThrPheTyrrIleValGlyLeuIleLeuSerMetGln 349
 DB 810 GGATCTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 869
 QY 309 roPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPhe 329
 DB 870 CTTTGTGGTCTTCCAGCTGCTCTTTCATCAGAGCATGCGCAGCTTTGGGGTCTTTG 929
 QY 329 lAlaLeuGlnAlaTyrrAlaPheLeuGlnTyrrLeuArgAspArgLeuThrLysGlnGlu 349
 DB 930 GTCTTCGCAGATTCATGCTTTTGGATTACCTGCGCAGCAAGTTGAATCCACAAT 989
 QY 349 hGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal 368
 DB 990 TTCAAGTCTTTTTCGGGAGCGTCTCTCTG---GTAGGCTTTTGTCTCTCTCTCACCG 1046
 QY 369 -IleTyrrLeuThrTyrrThrGlyTyrrIleAlaProTrpSerGlyArgPheTyrrSer 388
 DB 1047 GAGCTCTCTCATGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1106
 QY 388 rpAspThrGlyTyrrAlaIleHisIleProIleIleAlaSerValSerGluHisGlnP 408
 DB 1107 TGGATCCCTCTTATGCTAGAGAACACATCCCATCCTCTCTCTCTCTCTCTCTCTCT 1166
 QY 408 roThrThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaG 428
 DB 1167 CCACAACCTGGTCTCATACTATTTTACCTGCGAGCTCTCTCTCTCTCTCTCTCTCT 1226
 QY 428 lyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrrAla 448
 DB 1227 GCCTCTTACTGCTTTAGCAACCTGCTGATGCCGAGATTTTATCATCATGATGCTG 1286
 QY 448 lSerAlaValTyrrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 468
 DB 1287 TGACCGAGCTGCTTTTTCAGCTGTAATGGTGGCTTAATGCTAGTGTGGCAGCTGTTA 1346
 QY 468 alCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrrLeuGlyAsp 487
 DB 1347 TGTGCTTCTCTGTCATTTGAGTCTCCAGGCTCTCTCCACATACATGAAGATCTG 1406
 QY 487 spMetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArgAsn 507
 DB 1407 ACATAAGTCTGCTCA----- 1420
 QY 507 lGlyAsnLeuTyrrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLys 527
 DB 1421 -----GACAAG 1466
 QY 527 hrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeu 547
 DB 1467 ATGAA-----GTGGCAAGTGGGATGATGATGATGATGATGATGATGATGATGAT 1508
 QY 547 enMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrrSerSerProSer 567
 DB 1509 TCATACCTACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1568


```

PR 23-MAR-2000; 2000US-01916137P.
XX PR 11-JUL-2000; 2000US-00614150.
XX XX {PEKE } PE CORP NY.
XX PA Venter JC, Adams M, Li FWD, Myers EW;
PI PI
DR WPI: 2001-656860/75.
DR P-PDBE; ABB58692.
XX XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX PT
XX Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English.
XX CC
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABB55773-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX CC
XX SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 7,846-207 Length: 2855
Score: 2274.00 Matches: 435
Percent Similarity: 72.74% Conservative: 104
Best Local Similarity: 58.70% Mismatches: 176
Query Match: 51.67% Indels: 26
DB: 4 Gaps: 7
US-10-028-384-2 (1-826) x ABU02795 (1-2855)
QY 43 LysAlaLaAGlVGlyValAalProPolypsProAlaPro--AlaGlyLeuSerGlyGly 61
Db 380 AAGCGGCAGCGGGAAGCAAGAAATACGGCTGGCGCGCGTCCGCCAGGGA 439
QY 62 LeuSerGlnProAlaGlYrTrpGlnSerLeuLeuSerPheThrIleuPheLeuAlaTrp 81
Db 440 CTGTCTCACCTGGGACAACAGCAGCACCCTGGTGCAAGCTGGCCATTCTCATCTGGCAGG 499
QY 82 LeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGlu 101
Db 500 GTTTTATCATTTGCCACACGCTTGTCTCTGTCTGGATTCGAAAGCGTAATCATCATGAG 559
QY 102 PheAspProTrpPheAsnTyArgSerThrHisIleuLeuSerHisGlyPheTyGlu 121
Db 560 TTGGATTCGGTACTTCAATACCGCACCACCGGGTTCTCGCGGAGCAGGCGCTTTTACAG 619
QY 122 PheLeuAnsrTrpPheAspGluArgAlaTrpTyProLeuGlyValGlyGlyTrp 141
Db 620 TTCACAACCTGGTTGATGACCGGCGCTGTATCCCTCTGGCGCGCATCATCGCGCGCAC 679
QY 142 ValTyProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleuAnsrTrpLeuAn 161
Db 680 ATCTATCCCGGCTGATGCTCACTTCGGCGGCCCTGTACCGCTGATGGCTGTCTCAAT 739
QY 162 IleThrValHisIleArgAspValcysValPheLeuAlaProThrPheSerGlyLeuThr 181
Db 740 GTGACCATCGATACGGAAGGTGTGGCTCTTCTCTGGCGGCCCTTCTTCTCGCTGACC 799
QY 182 SerLieserThrPheLeuLeuThrArgGluLeuTrpPhanGlyValAglyLeuLeuAla 201
Db 800 AGCTGGTAGCTACGCCCTCACAAAGAGATACACAGCTGAGCTGAGCTGAGCTGGGCCC 859
QY 202 AlaCysPheIleAlaIleValProGlyTyrrIleSerArgSerValAlaGlySerPheAsp 221

```

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2004-204755/20.
 DR P-PSDB; ADL30662.
 XX
 XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 PT length human cDNAs.
 PS Example 1; SEQ ID NO 2694; 1340bp; English.
 XX
 XX This invention relates to a novel primers useful for synthesizing full
 CC length cDNA molecules that encode human proteins. Specifically, it refers
 CC to secretory or membrane proteins that are potential therapeutic agents/
 CC target molecules in the field of medicine, and in particular genes
 CC encoding proteins that are associated with signal transduction,
 CC glycoproteins and transcription. The present invention describes a method
 CC for efficiently cloning a full length human cDNA from both the 5' and 3'
 CC ends using the oligo-capping method. This polynucleotide sequence is a
 CC full length human cDNA clone of the invention.
 XX
 SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,28-216 Length: 1664
 Score: 2370.00 Matches: 443
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 53.85% Indels: 0
 DB: 12 Gaps: 0
 US-10-028-384-2 (1-826) x ADL30661 (1-1664)
 QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProtrp 20
 DB 28 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTGG 87
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 88 AGTGCCCTCATGGCCCTGGGAAACAGCCGCGAGCCAGCCAGCCGCGCGCGCGCGTGC 147
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 148 GGCACAAGGCGGCG 207
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 208 GGGCTGTCCAGCGCGCGTGGGTGGCGTGTCTCTCTCTCCATCTCTCTCTCTCTCTCT 267
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
 DB 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTGGCGTCTATCGCTTCGAAAGCATCATCCAC 327
 QY 101 GluPheAspProTrpPheAsnTyrArgSerThrHisLysLeuAlaSerHisGlyPheTyr 120
 DB 328 GAGTTCGACCGCGTGTGTTAACTATAGATCAACACATCATCTTGCATCTCTATGGGTTCTAT 387
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGly 140
 DB 388 GAATTTTAAATTTGGTTTGTATGAAGAGCATGGTATCCACTAGGAAGAATAGTAGGTGT 447
 QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
 DB 448 ACTGTTTACCAGGGGTTGATGATACCGCTGGCGCTTATTCTATGGATTTTAAATACATGG 507
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 508 AACATAACTGTTTCATATAGAGACGTGTGTGTGTCTCTTGCACCAACTTTTACGGCGCTT 567
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 DB 568 ACATCTATATCT 627

QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
 DB 628 GCTGCTTGTGTTTATTGCTATTGTACAGGCTACATATCTCGGTGAGTAGCTGGATCTTT 687
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSer 240
 DB 688 GATAAATAGGAGCATGCTATTGTCAGCTTCAGTTACATATCTATTATGGTAAATCT 747
 QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
 DB 748 GTAAAAACATGGGTGAGTTTGGCAATGTGCTGCTCTATCTATCTATCTATGTC 807
 QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 DB 808 TCTGCTTGGGTGGTATGATTATCATCAATCTATTCACATGCTATGTTGTTG 867
 QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
 DB 868 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTGTG 927
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 928 GGTATTATATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGAA 987
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlyTyrLeu 340
 DB 988 CACATGGCAGCTGCAGGTGCTTTGCTATGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1048 AGAGACCGATTAAACAAACAGAGTTCCAGACCTTTTCTTTTGGGTGATCACTAGCT 1107
 QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrp 380
 DB 1108 GCAGGTGCTGTGTTCTTGTATGCTATCTATTGCTTATACAGGTACATTCACCATG 1167
 QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIle 400
 DB 1168 AGTGGCGAGGTATTATTCATTGTTGGTATCGGTATGCAAAATACACATTCCAATTAT 1227
 QY 401 AlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHisIle 420
 DB 1228 GCATAGTGTCTGAGCATCAACCTAGCTTGGGTGCTTCTTCTTTGATCTACATATT 1287
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1288 CTTGTATGTACCTTCCAGCAGCGCTTGGTTCCTCATCAAAATATACACGATGAAGA 1347
 QY 441 ValPheValAla 444
 DB 1348 GTATTGGTGTCT 1359
 RESULT 8
 ID ABL02795 standard; cDNA; 2855 BP.
 XX AC ABL02795;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX OS pharmaceutical; gene; ss.
 XX FN Drosophila melanogaster.
 XX PD WO200171042-A2.
 XX PF 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.

XX The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,2e-216 Length: 1664
 Score: 2370.00 Matches: 443
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 53.85% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x AAK94164 (1-1664)

| | | | |
|----|-----|--|-----|
| QY | 1 | MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTTP | 20 |
| DB | 28 | ATGGCGGAGCCCTCGCCCGGAGAGGAGCAGCAGTCGTCCTCAACTCGTCCCGTGG | 87 |
| QY | 21 | SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys | 40 |
| DB | 88 | AGTGGCGCTCATGGCCCTGGGAAACAGCGCGGACCCACCGCGCGCGCGCGCGCGCG | 147 |
| QY | 41 | AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly | 60 |
| DB | 148 | GGGCACAGCG | 207 |
| QY | 61 | GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrLeuPheLeuAla | 80 |
| DB | 208 | GGGCTGTCTCGACCGCGCTGGGTGGCAGTCGGTCTCTCTCTCCATCCATCCCTCT | 267 |
| QY | 81 | TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis | 100 |
| DB | 268 | TGGCTTCGGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAGACATCCAC | 327 |
| QY | 101 | GluPheAspProTTPPheAsnTyrArgSerThrHisLysLeuAlaSerHisGlyPheTyr | 120 |
| DB | 328 | GAGTTCGACCCCTGGTTTAACTATATAGATCAACACATCATCTTCATCTCATGGT | 387 |
| QY | 121 | GluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly | 140 |
| DB | 388 | GAATTTTAAATGGTTTGGATGAAGACATGGTATCCACTAGAGAAATAGTAGTGGT | 447 |
| QY | 141 | ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu | 160 |
| DB | 448 | ACTGTTTACCAGGCTGATGAATACCGCTGGCTTATTCATTGGATTTTAAATATCAT | 507 |
| QY | 161 | AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu | 180 |
| DB | 508 | AACATAACTGTTCACATAGAGACGTGTGTGTGTCTTCTTCAACCACTTTTACGG | 567 |
| QY | 181 | ThrSerIleSerThrPheLeuLeuThrArgGlnLeuTrpAsnGlnIleAlaGlyLeuLeu | 200 |
| DB | 568 | ACATCTATATCTACTTCTCTCTTACAGAGAACTTTTGGAAACCAAGGAGGAGAC | 627 |
| QY | 201 | AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe | 220 |
| DB | 628 | GCTGCTGTTTATGCTATTGTACAGGCTACATATCTCGGTTCAGTAGTGCATCTT | 687 |
| QY | 221 | AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSer | 240 |
| DB | 688 | GATATGAAGGCACTGCTATTTTTCGACTTCAGTTTCACATACTATTATGGGTAAAT | 747 |

| | | | |
|----|--|--|------|
| QY | 241 | VallysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal | 260 |
| DB | 748 | GTAAAACTGGGTCAGTTTTTTGGACAATGTGCTGCTGCTTATCTATCTATCTAT | 807 |
| QY | 261 | SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu | 280 |
| DB | 808 | TCGCTTGGGCTGCTTATGCTATTTATCATCAATCTTATTCACATGCTATGTTG | 867 |
| QY | 281 | LeuLeuMetClnhArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrVal | 300 |
| DB | 868 | TTTACTGATGAGAGATACAGCAAGAGTCTACATAGCATATAGCACTTTCTACAT | 927 |
| QY | 301 | GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu | 320 |
| DB | 928 | GGTTTATATATTCAATGCAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAG | 987 |
| QY | 321 | HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu | 340 |
| DB | 988 | CACATGGCAGTGCAGGTGCTTTGCAATGTGCTGCAAGCTTATGCTTTCTTGCAG | 1047 |
| QY | 341 | ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla | 360 |
| DB | 1048 | AGAGACCGATTAAACAAACAGAGTCCAGACCCCTTTCTTTTGGGTGTATCACT | 1107 |
| QY | 361 | AlaGlyValAlaPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProT | 380 |
| DB | 1108 | GCAGTGTCTGTGTTCTTATGTCTATCTATTTGACTTATACAGTTTACATTCG | 1167 |
| QY | 381 | SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle | 400 |
| DB | 1168 | AGTGGCAGGTTTATTCATTGTGGGATACCTGGGTATGCAAAATACACATTC | 1227 |
| QY | 401 | AlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHisIle | 420 |
| DB | 1228 | GCATCAGTGTCTGAGCATCACTACAGCTTGGGTGCTTTCTTCTTGTATCTAC | 1287 |
| QY | 421 | LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu | 440 |
| DB | 1288 | CTGTATGTACCTTCCAGCAGCGCTTTGGTTCTGCTGCAATCAAAATATCAAC | 1347 |
| QY | 441 | ValPheValAla 444 | |
| DB | 1348 | GTATTTGGTGTCT 1359 | |
| AC | ADL30661; | | |
| DT | 20-MAY-2004 (first entry) | | |
| DE | Full length human cDNA clone SeqID 2694. | | |
| XX | human, medicine; signal transduction; glycoprotein; transcription; | | |
| XX | oligo-capping method; ss; gene. | | |
| XX | Homo sapiens. | | |
| XX | EP1396543-A2. | | |
| XX | 10-MAR-2004. | | |
| XX | 07-JUL-2000; 2003EP-00025638. | | |
| XX | 08-JUL-1999; 99JP-00194486. | | |
| XX | 11-JAN-2000; 2000JP-00118774. | | |
| XX | 02-MAY-2000; 2000JP-00183865. | | |
| XX | 07-JUL-2000; 2000EP-00114089. | | |
| XX | (REAS-) RES ASSOC BIOTECHNOLOGY. | | |
| XX | | | |

RESULT 7

ADL30661
 ID ADL30661 standard; cDNA; 1664 BP.

XX AC ADL30661;

XX DT 20-MAY-2004 (first entry)

XX DE Full length human cDNA clone SeqID 2694.

XX XX human, medicine; signal transduction; glycoprotein; transcription;
 oligo-capping method; ss; gene.

XX OS Homo sapiens.

XX XX EP1396543-A2.

XX XX 10-MAR-2004.

XX XX 07-JUL-2000; 2003EP-00025638.

XX XX 08-JUL-1999; 99JP-00194486.

XX XX 11-JAN-2000; 2000JP-00118774.

XX XX 02-MAY-2000; 2000JP-00183865.

XX XX 07-JUL-2000; 2000EP-00114089.

XX XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX XX

QY 717 MetTyrLysMetSerTyrTyrArgPheGlyCysMetGlnLeuAspPheArgThrProPro 736
 Db 2084 ARGTAACAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTGGACTACAGAGTCCATCT 2143
 QY 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
 Db 2144 GGATATGATCGCACACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG 2203
 QY 757 GluGluAlaPheThrSerGluHisTyrPheValArgIleTyrLysValLysAlaProAsp 776
 Db 2204 GAGGAGGCGCTACACACAGAACTGCTTGTTCGCACTATAGGGTGAAGAGCGCAT 2263
 QY 777 -----AsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGln 794
 Db 2264 GAGTTCAATAGACCACTCACTCAAGACCAAGGAGAGAAGC-----ATTCTCCAGCA 2314
 QY 795 LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814
 Db 2315 AACTTCATTTGAGAAAGAACTCTAAGCGTCGCAAGGCTACATACAAACCGCGTT 2374
 QY 815 PheLysLysGlyLysLysLysSerLysLysThr 825
 Db 2375 GTTGTAAAGGAAACCAACCACTTGAATAAACC 2407

RESULT 5

ABL13246/c

ID ABL13246 standard; cDNA; 4922 BP.

XX ABL13246;
 XX ABL13246;XX 26-MAR-2002 (first entry)
 XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.OS Drosophila melanogaster.
 OS Drosophila melanogaster.XX WO200171042-A2.
 XX WO200171042-A2.XX 27-SEP-2001.
 XX 27-SEP-2001.XX 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2001; 2001WO-US009231.XX 23-MAR-2000; 2000US-0191637P.
 XX 23-MAR-2000; 2000US-0191637P.XX 11-JUL-2000; 2000US-00614150.
 XX 11-JUL-2000; 2000US-00614150.XX (PEKE) PE CORP NY.
 XX (PEKE) PE CORP NY.XX Venter JC, Adams M, Li PWD, Myers EW;
 XX Venter JC, Adams M, Li PWD, Myers EW;XX WPI; 2001-65860/75.
 XX WPI; 2001-65860/75.XX P-PSDB; ABB69143.
 XX P-PSDB; ABB69143.PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT Genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PT interactions.XX Claim 1; SEQ ID NO 34220; 21bp + Sequence Listing; English.
 XX Claim 1; SEQ ID NO 34220; 21bp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB15175), and the encoded proteins (AB557737-
 CC AB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
 XX Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;

Alignment Scores:

Pred. NO.: 1.8e-260 Length: 4922
 Score: 2838.00 Matches: 557
 Percent Similarity: 76.24% Conservative: 88
 Best Local Similarity: 65.84% Mismatches: 114
 Query Match: 64.49% Indels: 89
 DB: 4 Gaps: 10

US-10-028-384-2 (1-826) x ABL13246 (1-4922)

QY 63 SerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 Db 3816 AGCAAGGTGGTGGCTACAGACGCTTAATCACTTCGCCATCTGCTAATGCGCTGGCTG 3757
 QY 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 Db 3756 GCGGATTTTCTCTCGCTCTTCGCGGTCATCCGTTTCGATCGCATATTCATGAGTT 3697
 QY 103 AspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe 122
 Db 3696 GATCCGTGGTTCAACTACCGGGCCACCGCTACATGTCAGAAATGGTTGGTACAACTTC 3637
 QY 123 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrVal 142
 Db 3636 CTAACTGGTTCGACGAGCGCATGGTAATCCGCTCGGAGGATTTGGGGGTGATCCGTC 3577
 QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThrIleAsnIle 162
 Db 3576 TATCCCGGCTGATGATTACGTCGCGGGAATCCATTCGCTCGCTACGCTACTCAACATA 3517
 QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 3516 CCGGTTCATATTCGTGACATCTCGGTTCCTGGCGCGCATCTTCAGTGGCTGACCTCC 3457
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuAlaAla 202
 Db 3456 ATCTCACCTACCTGCTGACCAAGGAGCTGTGGTCCGCGGCGCGGCTCTTCGCCGCC 3397
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 Db 3396 AGTTTCATCGCCATCGTGGCTGCTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGT 3337
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLys 242
 Db 3336 GAGGCGATTGCCATATTCGCCCTGCAGTTCACCTACTTCTCTGGGTGGTGGTGGTGGT 3277
 QY 243 ThrGlySerValPheThrThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 Db 3276 ACTGGATCCGTGTTCTGGTGGCGCGCGCGCTTTGTCTCTACTTCTACATGGTGGCTG 3217
 QY 263 TrpGlyGlyTyrValPheIleIleLeuLeuIleProLeuHisValPheValLeuLeuLeu 282
 Db 3216 TGGGGTGGGTACGTTGTTTCATCACTCAACCTGTATACCCCTGACGCTTCTGCTACTT 3157
 QY 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 Db 3156 ATGGCAGGTATCTCGCGGCTGTGTCGACCATACGACCTTCTACATCTGAGGACTG 3097
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 3096 CTGTTCTCCATGACAGATCCCTTCGTGGGATTTCCAAACCGGATACGACGAGTGAACATG 3037
 QY 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 Db 3036 GCTGCGCTGGAGTGTGTGTCTCTATGCGCGGTGGCCACCTTGGCGCATTTGAGTCC 2977
 QY 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaGly 362
 Db 2976 GTGCTGTCCGCAACGAGTTCGGAAGTGTTCATCTCGTGGCGGATTCGCTGGCGCTG 2917
 QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGly 382

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,35e-271 Length: 2699
 Score: 2950.00 Matches: 557
 Percent Similarity: 83.79% Conservative: 89
 Best Local Similarity: 72.24% Mismatches: 113
 Query Match: 67.03% Indels: 12
 DB: 4 Gaps: 7

US-10-028-384-2 (1-826) x ABL13247 (1-2699)

QY 63 SerGlnProAlaGlyTyrPrlGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 DB 107 AGCAAGTGGTGGCTACAGAGCCTAATCACCTTCGCCATCTCTGTAATCCCTGGGTG 166
 QY 83 AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 DB 167 GCGGATTTTCCTCTCGCCCTTCGCGGTCTATCCGTTTCGAGTCGATTATCCATGAGTTT 226
 QY 103 AspProTrpPheAntyArgSerThrHisIleLeuAlaSerHisGlyPheTyrGluPhe 122
 DB 227 GATCCGTGGTTCAACTACCGGGCCACCGCTACATGGTCAGAGTAATGTTGGTACAACTTC 286
 QY 123 LeuAsnTrpPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyThrVal 142
 DB 287 CTCACCTGGTTCGAGAGCGCGCATGATCCGCTCGGCAGGATGTTGGGCGGTACCGTTC 346
 QY 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 DB 347 TATCCCGCCCTGATGATAGTTCGCGGGAATCCATTTGCTGCTGCACATCTCAACATA 406
 QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 DB 407 CCGGTCCATATTCGTGACATCTCGTGTCTCGCCCGATCTTCAGTGGCTCGACCTCC 466
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
 DB 467 ATCTCCACCTACCTGCTGACCAAGAGCTGTGCTCGCGGGCGCGCCCTCTTCGCGGCC 526
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 DB 527 AGCTTCATCCGTCATCGTGGCTACATCAGTAGTCTGGTGGCTGGATCGTACGATAAC 586
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSerValIys 242
 DB 587 GAGGGCATTTGCCATATTCGCGCTGCACTTCCATCTCTGCTGGGTGGCTCAGTGAAG 646
 QY 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 DB 647 ACTGGATCCGTTCTGCTGGCGCGCAGCGCTTTGTCTTCTTCTACATCTCTGGGCTCC 706
 QY 263 TrpGlyGlyTyrValPheIleAlaLeuIleAsnLeuIleProLeuHisValPheValLeuLeu 282
 DB 707 TGGGGTGGCTACGTTTTCATCATCACTGATACCCCTGACGCTCTCTGCTGCTGCTCAT 766
 QY 283 MetGlnArgTyrSerIysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 DB 767 ATGGCAGGTACTCGCGCGTCTGCTGACCGAGCTACAGCACCTTCTTACATCTCTGGGCTG 826
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 DB 827 CTGTTCTCATGCGAGATCCCTCTCGTGGGATTCACCGGATCGACCGACGATGACACATG 886
 QY 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 DB 887 GCTGCGCTGGGAGTGTGTGCTCTTATGGCCGTGGCCACCTTGGCCATTTGCGATCC 946
 QY 343 ArgLeuThrLysGlnPheGlnThrLeuPhePheLeuGlyValSerLeuAlaGly 362
 DB 947 GTGCTGTCCGCAACGAGTTCCGGAAGCTGTTCATCTGCGCGGATGTTGTTGGGCGT 1006

QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGly 382
 DB 1007 GCGGTCTTTTGGCGGTGGTGTCTCACCATGTCTGGCGGTGGTGGCGGTGGAGTGA 1066
 QY 383 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaIysIleHisIleProIleIleAlaSer 402
 DB 1067 CGCTTCTACTCTGCTGGGATCTGGCTACGCCAAGATCCACATTCCTCATCTTCATCC 1126
 QY 403 ValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIleLeuVal 422
 DB 1127 GTGTGCGAGCATCAGCCACCATCTGTTCTCTGTTCTCTGATCTGCACTCTCGGTG 1186
 QY 423 CysThrPheProAlaGlyLeuTrpPheCysIleIysAsnIleAsnAspGluArgValPhe 442
 DB 1187 TGGCGCTTCCAGTGGGAGTGTGTTACTTCATCAAGCAGATCAACAGAGCGGCTTTC 1246
 QY 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
 DB 1247 GTGTGCTGTACGCCATCAGTGGGTTCCTGCTGCTGATGTTGATGTTGATGTTG 1306
 QY 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
 DB 1307 ACCCTCACCGCGTGGTGTGCTGCTGGCGAGTGGCTTTCGGGACTGTTGGATGTG 1366
 QY 483 TyrLeuGlyAspAsp---MetIysArgGluAsnProValGluAspSerSerAspGlu 501
 DB 1367 TTCCTGCAAGAGATTCGTTCTAAGCGAATGGGCAAGCCTAAGCGGAGCCACCGAAGT 1426
 QY 502 Asp-----AspIysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArg 518
 DB 1427 GATGAAGCTCAGATTCATTCAGAGAAGAGCGCTGTACGACAGGCTGGCAGCTG--- 1483
 QY 519 LysHisAlaThrGlnGlnGlnLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
 DB 1484 AGCATCGTACTAAGCATGATGCCCGAGGAGTATGCGCGTCAGCTCCAACTCCAAAGAT 1543
 QY 538 IleValThrMetLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 557
 DB 1544 ATTGTTATTTTGGCGGTCTAATGCTGTGTGATGTTGCTGCTGCTGCTGCTGCTGCTG 1603
 QY 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
 DB 1604 ACCAGCAATGCTTACTCCAGTCCCTTCTGCTTGGCTTCCACACAGCTCAAGATGGA 1663
 QY 577 ThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAsp 596
 DB 1664 TCCCGCAACATTTTAGACGATTTACAGAGGCTTACTTACTGGCTTTCGACAGACACTGCC 1723
 QY 597 GluHisAlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 616
 DB 1724 GATGATCTCGGCTTATGCTTGGTGGATTCAGGATACCGATAGCGGAATGGCAAC 1783
 QY 617 ArgThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 636
 DB 1784 AGAACGACCTAGTGGATTAATACCTGGAGAACATAGTACATAGCGCTGTTGGCAAG 1843
 QY 637 AlaMetSerSerAsnGluThrAlaAlaTyrIysIleMetArgThrLeuAspValAspTyr 656
 DB 1844 GCAATGCTTCAACCGAGGAGAGATCTTACGAAATATATGACATCTCTTACGCTGACTAC 1903
 QY 657 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
 DB 1904 GTTTTGGTGTATCTTTGGCGGTGTGATCGCTATCTGCGGATGATATCAACAAGTTCCTG 1963
 QY 677 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
 DB 1964 TGGATGCTCCGAATTTGCTCAGGAGAGATCTCCCAAGGACATTAAGGAAAGCGATTTACT 2023
 QY 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716
 DB 2024 ACCGACCGGCTGATTTCAAGGATAGATCCGAGAGTGTCTCGGCGCTCTCACTGCTT 2083

Qy 760 PheThrSerGluHieTrLeuValArgIleTyrlsValysAlaProAspAsnArgGlu 779
 Db 2345 TTTACATCAGACCACTGGCTTGTGAGGATATATAAGTGAAGAACCTGACACAGGGAG 2404
 Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrlsSerLys 799
 Db 2405 ACATAGGTCAAAACCTCGAGTACCAACATGTCCTCCCAACAGAGATTTGTCAAG 2464
 Qy 800 LysThrThrLysArgLysArgGlyTyrlsLysAsnLysLeuValPheLysLysGlyLys 819
 Db 2465 AAGACTACTAAAGGAAGCGTGGCTACCTGTTAAAAATAAGTGTGTTTGAAGAAGCAAG 2524
 Qy 820 LysIleSerLysLysThrVal 826
 Db 2525 AAGACCTCTAAGAAAGACTGTT 2545

RESULT 3

ADD94789
 ID ADD94789 standard; DNA; 2417 BP.

XX AC ADD94789;

XX 29-JAN-2004 (first entry)

XX Drosophila melanogaster STT3 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; Gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide;
 KW fruit fly; gene; ds; STT3.

XX Drosophila melanogaster.

XX WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX Perreault C, McBride K;

XX WPI; 2003-559122/52.

XX P-FSDB; ADD94790.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.

XX Claim 6; SEQ ID NO 7; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response

CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the Drosophila melanogaster STT3 gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.

XX Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.14e-271 Length: 2417
 Score: 2950.00 Matches: 557
 Percent Similarity: 83.79% Conservative: 89
 Best Local Similarity: 72.24% Mismatches: 113
 Query Match: 67.03% Indels: 12
 DB: 10 Gaps: 7

US-10-028-384-2 (1-826) x ADD94789 (1-2417)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
 Db 98 AGCAAGGTGGCTGGCTTACAGCAGCCTAATCACCTTGGCATCTCTGCTTAATCGCGCTGCTG 157
 Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
 Db 158 GCCGATTTTCCTCTCGCCTCTTCGCGTATCTGCTTCGAGTCGATATCCATGAGTTT 217
 Qy 103 AspProTrpPheAsnTyArgSerThrHisHisLeuAlaSerHisGlyPheTyrlsGluPhe 122
 Db 218 GATCCCGTGGTTCACCTTACCGGCCACCGCTACATGGTGCAGATGGTGTGTACCACTTC 277
 Qy 123 LeuAsnTrpPheAsnGluArgAlaTrpTyrlsProLeuGlyArgIleValGlyGlyThrVal 142
 Db 278 CTCACTGGTTCAGCAGCGCATGGTATCCGCTCGCAGCATGTGGCGGTGATCCGTC 337
 Qy 143 TyrProGlyIleuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 Db 338 TATCCCGGCTGATGATTACGTCGCGCGGAATCCATTGGCTGTGTGCACGACTCAACATA 397
 Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 398 CCGGTCCATATTCGTGACATCTGCTGTCTTCTGGCGCGATCTTCAGTGGCTGACCTCC 457
 Qy 183 IleserThrPheLeuLeuThrArgGluLeuTrpAsnGlyAlaGlyLeuLeuAlaAla 202
 Db 458 ATCTCCACCTACCTGCTGACCAAGAGCTGTGTGTGTGGCGCGCGCGCTCTTCGCGGCC 517
 Qy 203 CysPheIleAlaIleValProGlyTyrlsSerArgSerValAlaGlySerPheAspAsn 222
 Db 518 AGCTTCATGCGCATGCTGCTGGCTACATCATAGTAGTGGTGGTGGATGTAGGATAAC 577
 Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrlsLeuTrpValLysSerValLys 242
 Db 578 GAGGGGATTTGCCATATTCGCGCTGACCTTCCCTTCTGTGGTGGTGGCTCAGTGAAG 637
 Qy 243 ThrGlySerValPheThrTrpMetCysCysLeuSerTyrlsPheTyrlsMetValSerAla 262
 Db 638 ACTGGATCCGTGTTCTGTGGCGCGCGCGCTTGTCTTCTTCTATCATGTGTGCGCC 697
 Qy 263 TrpGlyGlyTyrlsValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeu 282
 Db 698 TGGGGTGGTTCATGTTTATCATCACTGATACCCCTGACGCTTCCTGCTGCTGCTCAT 757
 Qy 283 MetGlnArgTyrlsSerLysArgValTyrlsAlaTyrlsSerThrPheTyrlsIleValGlyLeu 302
 Db 758 ATGGGAGGTACTTCGCGCGCTGCTGTGACCGACCTACAGCACCTTCTATCATCTGGGACTG 817
 Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 818 CTGTTCTCCATGAGATCCCTTCTGTGGGATTCACAGATACGACACGACGACACATG 877
 Qy 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrlsAlaPheLeuIleTyrlsLeuAsp 342
 Db 878 GCTGCGCTGGAGTGTGTTGTGCTCTCTTATGGCGCGCGCACCTTGGCGCCATTTGCGACTCC 937

Db 137 AGCGGCTCATGGCTCTGGGGAACAGCGCGCACCGGACCATGGGCGCGGAACCCAGAGC 136
Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProlyProAla---ProAlaGlyLeuSer 59
Db 197 GCGTCCAGGCGGCGGCGG-----CGGAAGCGGCGGCGGCGGCGGCGGCGGCTGCC 244
Qy 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeu 79
Db 245 GCGGGCTTGTGGAGCGCGCGCGGGTGGCAGTCTGTCTCTCTCAACAACCTCTTCCTG 304
Qy 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
Db 305 GCGTGGCTGGCGGCTTACGCTCGGCGCTCTTCGCGGCTATCCGCTTCGAGAGCATATC 364
Qy 100 HisGluPheAspProTrpPheAsnTyArgSerThrHisHisLeuAlaSerHisGlyPhe 119
Db 365 CAGGAGTTCGACCGCGGTGTTAACTATAGATCAACACATCATCTTGATCTCATGATTC 424
Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
Db 425 TATGAGTCTTCTAAATTTGGTTTGTGAAAGAGCATGGTACCACCTGGGAAGATAGTGGGT 484
Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 485 GGCACCGTTTACCAGGGTGTGATTAACAGCTGCGCTTATTCATTCGATTTTAAATACA 544
Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 545 TTGAACATACAGTTCACATAGAGATGGTGTGATTCCTTGGACCAACTTTTAGCGGC 604
Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 605 CTTACATCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAACCAAGAGGAGGAGCTT 664
Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
Db 665 CTAGCTGCTGCTTCACTGCTATCGTACAGGATACATATCTCGGTGAGTGGCGGATCC 724
Qy 220 PheAspAsnGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIys 239
Db 725 TTTGATAAATAGAGGCAATGCCATTTTGGCGTTCACTTCACTTACTTATGGGTAAAG 784
Qy 240 SerValIleThrCysSerValPheThrPheThrMetCysCysLeuSerTyrPheTyrMet 259
Db 785 TCTGTGAAGCGGCTGCTGTCTGCAATAGTCTGCTCTGCTGCTATTCATATTCATATG 844
Qy 260 ValSerAlaTrpGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
Db 845 GTCTCTGCGTGGGAGGTATGTGTTTCATCATCAACCTCATCCCTCTCCATGTGTGTG 904
Qy 280 LeuLeuMetGlnArgTyrSerIysArgValTyrIleAlaTyrSerThrPheTyrIle 299
Db 905 TTGCTGCTGATGCAGAGGTACAGCAGAGAGGTCTACATAGCATATAGCATTTGTACATT 964
Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db 965 GTGGGTATTAATATATCCATGAGATACCTTTTGGGATTTACGCCAATCAGAACAAAGC 1024
Qy 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
Db 1025 GAGCATATGGCAGCTCAGGTGCTTGGCGTGTGCGAGGTCTTACGCTTTTTCGAGTAT 1084
Qy 340 LeuArgAspArgLeuThrIysGlnPheGlnThrLeuPhePheLeuGlyValSerLeu 359
Db 1085 CTGAGAGACCGGTGCAAAACAGGAGTTCCAGACCCCTTTTCTTTTGGGTGTCTCACTA 1144
Qy 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
Db 1145 GCTGCGGCGCTGTCTCTTAGTGTCTATCTATCTGACATACACAGGTATATGACACCA 1204
Qy 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
Db 1205 TGGAGTGGCAGGTTTATTTACTATGGGATCTGGGTATGCAAAATAACACATTCACATT 1264

Qy 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheAspLeuHis 419
Db 1265 ATTGCATCAGTGTCTGAACATCAGCCTAGCATAGGTGTCTTTCTTTGATCTACAT 1324
Qy 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLeuAsnIleAsnAspGlu 439
Db 1325 ATTTCTGTATGTACCTTCCAGCAGCGCTATGGTCTGCATCAAAAATATCAACGATGAA 1384
Qy 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
Db 1385 AGAGTATTTGTGCTCTGTATGCGATCAGTGTGTGTACTTTGCCGAGAGTATGTTGGG 1444
Qy 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
Db 1445 CTGATGCTGACTCTGACCCCGGTCTGCTGCATGCTGTGGGCCATCGCTCTTCCCATGTT 1504
Qy 480 PheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSer 499
Db 1505 TTTGAGCAGTATTTGGGGGATGACATGAAAAGGGAAAAACCCACCTGTGGAGACAGCAGT 1564
Qy 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
Db 1565 GATGAGGATGACAAAAGAAACCCAGGAAACTTGTATGACAAAGCAGGTAAAGTGAGGAAG 1624
Qy 520 HisAlaThrGlnGlnGlnIlyThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
Db 1625 CATGTGACAGCAGCAGAGAAACCTGAAGAGGGGCTTGGGCCCCCAACATCAAAAGCATTGTG 1684
Qy 540 ThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSer 559
Db 1685 ACCATGCTGATGCTCATGCTCTCTGATGATGTTTCGGGTCCACTGCACGTGGGTCAACAAGC 1744
Qy 560 AsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579
Db 1745 AACCCCTACTCCAGTCCAGTGGTCTTCTCTCTACATCATGATGTGTACAGGAAT 1804
Qy 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599
Db 1805 ATATTAGATGATTTAGAGAAGCGTACTTTTGGCTGAGACAAAACACGATGAACACGCC 1864
Qy 600 ArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
Db 1865 CGGTCTATGCTGTGTGGGACTACGCTATCAGATGCTGTGGCATGGCCACAGACCACT 1924
Qy 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
Db 1925 CTGGTGGATACAAACACCTGGAAACACAGCCACATCGCAGCTGTGGTGGGAAAGCTATGTCT 1984
Qy 640 SerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVal 659
Db 1985 TCCATGAAACCGCGGCTATATAAATCATGAGGTCCCTTGATGTGATTTATGTGTGGTT 2044
Qy 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
Db 2045 ATTTTCGAGGAGTGTATGGCTATTCGGGGGACGATATCAACAAGTTCCTCTGGATGTC 2104
Qy 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
Db 2105 AGGATAGCTGAAGGGAGCATCCCAAGACATCCGGGAAGGTGACTATTTTACCACGAGCAG 2164
Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrIys 719
Db 2165 GGAGAGTTCGAGTAGCAAAAGCTGGGTCTCTCTACTCTCTTAACTGGCTTATGTATATAA 2224
Qy 720 MetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAsp 739
Db 2225 ATGTCATACATACAGTTTGGAGAAATGACAGTAGATTTTCGCACCTCCCGAGCTTTGAC 2284
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheIleHisLeuGluAla 759
Db 2285 CGAACACGTAACTGCTGAGATTGGAAATTAAGACATTAATAATTCAAGCATTTGGAGGAAGCT 2344

Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCGAGTAAAGTGAGGAAACAT 1560
 Qy 521 AlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThr 540
 Db 1561 GCNACTGACAGGAAAGAACTGAAGAGGATAGGCCCTAATATAAAGACATGTGCACC 1620
 Qy 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 Db 1621 ATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 Qy 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrAsnIle 580
 Db 1681 GCCTACTCTAGTCCAGGTAGTCTCTGCTCATACATCATGATGCGACCAAGATATC 1740
 Qy 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 Db 1741 TTAGATGATGATTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA 1800
 Qy 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 Db 1801 GTAATGCTTGGTGGGATATGGCTATCAGATAGTGGATGGCTATAGACTAGCTTGG 1860
 Qy 621 ValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 Db 1861 GTGGATAATAACACCTGGAAATACAGCCACATAGCCTGTGGGAAAGTATGTCTTCT 1920
 Qy 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
 Db 1921 AATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGATGATATGTTTGGTTAT 1980
 Qy 661 PheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArg 680
 Db 1981 TTTGGAGGGGTATTTGGCTATTTCTGGTATGATATCAACAATTTCTCTGGATGGTAGG 2040
 Qy 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 2041 ATAGCTGAAGGAGAAACATCCAAAGACATTCGGGAAGTGACTATTTATCCCCACAGGA 2100
 Qy 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 2101 GAATTCGGGTAGACAAAGCAGGATCCCTCTCTTTGTGATTCCTCTTATGATAAAATG 2160
 Qy 721 SerTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
 Db 2161 TCATACTACAGATTGGAGAAATGCGCTGGATTTCTGACACCCCGAGTTTGCACCGA 2220
 Qy 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 Db 2221 ACACGTAATGCTGAGATGGAATAGGACATTAATTCAAACATTTCGGAAGAGCCTTT 2280
 Qy 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2281 ACATCAGACACTGGCTTGTAGGATATATAAGTAAAGCACCTGATACAGGAGACA 2340
 Qy 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2341 TTAGATCAGAACCTCGAGTCCCAACATTTTCCAAACACAGAAATTTTGTCAAGAG 2400
 Qy 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysGlyLysLys 820
 Db 2401 ACTACCAAGAGGAGCGTGGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAA 2460
 Qy 821 IleSerLysLysThrVal 826
 Db 2461 ATATCTAAGAGACTGTT 2478

RESULT 2

ADD94785
 ID ADD94785 standard; cDNA; 2710 BP.
 XX
 AC ADD94785;
 XX
 DT 29-JAN-2004 (first entry)

XX Mouse SIMP cDNA sequence.
 DE source of immunodominant MHC-associated peptide; SIMP; MHC;
 XX major histocompatibility complex; human leukocyte antigen; HLA;
 KW cycostatic; immunosuppressive; antitumor therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ss.
 XX
 OS Mus musculus.
 XX
 XX WO2003054008-A2.
 PN 03-JUL-2003.
 XX
 XX 18-DEC-2002; 2002WO-CA001967.
 PF 20-DEC-2001; 2001US-00028384.
 PR
 XX (COMP-) COMPATIGENE, INC.
 PA Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 DR P-PSDB; ADD94786.
 XX
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Disclosure; SEQ ID NO 3; 66pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antitumor therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the mouse SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2710
 Score: 4263.50 Matches: 805
 Percent Similarity: 97.70% Conservative: 3
 Best Local Similarity: 97.34% Mismatches: 14
 Query Match: 96.88% Indels: 5
 DB: 10 Gaps: 2

US-10-028-384-2 (1-926) x ADD94785 (1-2710)

Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 Db 77 ATGGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTGGTCCCTCAACTGCTCCCGTGG 136
 Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40

XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX DR P-PSDB; ADD94784.
 XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX PS Claim 6: SEQ ID NO 1; 65pp; English.
 XX CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2481
 Score: 4401.00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x ADD94783 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 DB 1 ATGGCGGAGCCCTGGCCCGCGAGAGCAAGCAAGCGTCCCTCAACTCGTCCCGTG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGCCTCATGGCCCTGGGAARAGCGGCGAGCGGCGCCACCGCGGCGGCGCCAGTGC 120
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 GGCACAAAGCGGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrLeuPheLeuAla 80
 DB 181 GGGCTGTGCGAGCGCGCGTGGGTGGCGTGGTCTCTCTCTCCACCACTCTCTCTCTCT 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleLeuHis 100
 DB 241 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGTCTATCGCTTCGAAAGCATATCCAC 300
 QY 101 GluPheAspProTrpPheAsnTyArgSerThrHisIleuAlaSerHisGlyPheTy 120
 DB 301 GAGTTCGACCCCGTGTTCATATAGATCAACATCATCTTCATCTCATGGGTTCAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATGGTTTGTATGAAGAGCATGGTATCCATAGAGAGATAGTAGTGT 420
 QY 141 ThrValTyProGlyLeuMetIleThrAlaGlyLeuIleHisTrpLeuAsnThrLeu 160

DB 421 ACTGTTTACCAGGGTTGATGATAACCGCTGCTTATTCATTGGATTAAATACATTG 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 481 AACATAACTGTTTACATATAGAGAGCGTATGTGTGTTCTTGCACCAACTTTAGCGGCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 DB 541 ACATCTATATCTCTCTCTCTTACAGAGAACCTTTGGAAACCAAGGAGGAGGAGCTTTA 600
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyTrpIleSerArgSerValAlaGlySerPhe 220
 DB 601 GCTGCTGTTTATTTGCTATTTACAGGCTACATATCTCGCTCAGTAGTGCATCTT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyTrpIleTrpVallySer 240
 DB 661 GATAAATGAAGGCACTGCTATTTTGCACCTTCAGTTTACATCTATTTATGGGTAAATCT 720
 QY 241 ValIysThrGlySerValPheThrMetCysCysValSerCysLeuSerTyPheTrpMetVal 260
 DB 721 GTAAAACTGGGTCAGTTTGGACAATGCTGCTCTTATCTCTATCTATATGCTC 780
 QY 261 SerAlaTrpGlyGlyTyValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 DB 781 TCTGCTTGGGCTGGTATGTTTATCATCAATCTTATTCACCTCATGTATTTGTGTG 840
 QY 281 LeuLeuMetGlnArgTySerLysArgValTyTrpIleAlaTySerThrPheTyTrpVal 300
 DB 841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTCTTACATTTG 900
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTTTAAATATTAATATGAGATACCTTTTGGGGAATCCAGCAATCAGAAAGTGA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyAlaPheLeuGlnTyLeu 340
 DB 961 CACATGSCAGCTCAGGCTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTGAGTATCTG 1020
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGACCGATTAAACAAACAAAGAGTTCCAGAGCCCTTTCTTTGGGTATACATAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValIleTyLeuThrTyThrGlyTyTrpIleAlaProTrp 380
 DB 1081 GCAGTCTGCTGTTCTTCTAGTGTCACTATTTGACTTATACAGGTTACATTGCACCATG 1140
 QY 381 SerGlyArgPheTySerLeuTrpAspThrGlyTyAlaLeuIleHisIleProIleIle 400
 DB 1141 AGTGGCAGGTTTTTATTCATTGTTGGGATCTGGGTATGCAAAATACATTCACATTTAT 1200
 QY 401 AlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheAspLeuHisIle 420
 DB 1201 GCATCAGTGTCTGAGCATCAACCTAGCTGCTGGGTGCTTTCTTCTTGTATCTACAT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLeuAsnIleAsnAspGluArg 440
 DB 1261 CTGTATGTACCTTCCAGCAGGCGCTTTGGTCTCTGCTATGCAAAATATCAACGATGAA 1320
 QY 441 ValPheValAlaLeuTyAlaIleSerAlaValTyPheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTTGCTCTATATGCAATCAGTGTGCTGCTACTTTGCTGGAGTGATGCTGGCGCTG 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 DB 1381 ATGTGTACTTTGACTCCAGTCGTGTATGCTGTGCTGCAATTCGCTTTTCAATGTTTTT 1440
 QY 481 GluHisTyTrpLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCACTATTGGCGGATGACATGAAAGGGAATCCACCTGTGCGAGCAGCAGCTGAT 1500
 QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTyAspLysAlaGlyValArgLysHis 520

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 ; Search time 951.963 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 4401

Sequence: 1 MAEPSAPESKHKSLSNPPW.....GVINKLVFKGKIGSKTKTV 826

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US10028384/runat.14122004.131518.8007/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -QMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEXT=ptc -NOEM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 -CGEN.1.1.1986 -runat.14122004.131518.8007 -NCRU=6 -ICPU=3
-NO-MAP -LARGQUERY -NEG_SCORE=50 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N Geneseq 23Sep04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|----------|---------------------|
| 1 | 4401 | 100.0 | 2481 | 10 | ADD94783 | Add94783 Human SIM |
| 2 | 4263.5 | 96.9 | 2710 | 10 | ADD94785 | Add94785 Mouse SIM |
| 3 | 2950 | 67.0 | 2417 | 10 | ADD94789 | Add94789 Drosophila |
| 4 | 2950 | 67.0 | 2699 | 4 | ABL13247 | ABL13247 Drosophila |
| 5 | 2938 | 64.5 | 4922 | 4 | ABL13246 | ABL13246 Drosophila |
| 6 | 2370 | 53.9 | 1664 | 4 | AAK94164 | AAK94164 Human ful |

| | | | | | | |
|----|--------|------|------|----|-----------|---------------------|
| 7 | 2370 | 53.9 | 1664 | 12 | ADL130661 | Adl130661 Full leng |
| 8 | 2274 | 51.7 | 2855 | 4 | ABL02795 | ABL02795 Drosophila |
| 9 | 2266.5 | 51.5 | 3094 | 10 | ADD94791 | Add94791 Mouse ITM |
| 10 | 2250 | 51.1 | 2472 | 10 | ADD94793 | Add94793 Human ITM |
| 11 | 2250 | 51.1 | 2472 | 10 | ADH28844 | Adh28844 Human chr |
| 12 | 2236 | 50.8 | 2760 | 5 | ABV24502 | Abv24502 Human pro |
| 13 | 2181.5 | 49.6 | 2466 | 10 | ADB69900 | Adb69900 C. neofor |
| 14 | 2179.5 | 49.5 | 2332 | 8 | ABT20789 | Abt20789 Aspergill |
| 15 | 2164 | 49.2 | 2256 | 6 | ABZ32036 | Abz32036 Candida a |
| 16 | 2023 | 46.0 | 2803 | 8 | ABT20191 | Abt20191 Aspergill |
| 17 | 2023 | 46.0 | 3969 | 8 | ABT17781 | Abt17781 Aspergill |
| 18 | 2023 | 46.0 | 4603 | 8 | ABT19595 | Abt19595 Aspergill |
| 19 | 2021.5 | 45.9 | 2733 | 10 | ADD94787 | Add94787 Yeast STT |
| 20 | 2021.5 | 45.9 | 6153 | 4 | ABL02794 | ABL02794 Drosophila |
| 21 | 1958.5 | 44.5 | 2000 | 10 | ACC61293 | Acc61293 Gene sequ |
| 22 | 1958.5 | 44.5 | 2000 | 10 | ADK63659 | Adk63659 Disease t |
| 23 | 1941.5 | 44.1 | 4738 | 10 | ADB69178 | Adb69178 C. neofor |
| 24 | 1939.5 | 44.1 | 2882 | 10 | ADB69539 | Adb69539 C. neofor |
| 25 | 1897 | 43.1 | 1948 | 8 | ABT18969 | Abt18969 Aspergill |
| 26 | 1896.5 | 43.1 | 1543 | 2 | AAK85055 | Aak85055 Human sec |
| 27 | 1896.5 | 43.1 | 1543 | 8 | ACD18981 | Acid18981 Novel hum |
| 28 | 1896.5 | 43.1 | 1543 | 12 | ADG78372 | Adg78372 Human sec |
| 29 | 1896.5 | 43.1 | 1543 | 12 | ADN60663 | Adn60663 Human sec |
| 30 | 1894.5 | 43.0 | 1209 | 4 | AAH33264 | Aah33264 Human col |
| 31 | 1894.5 | 43.0 | 1209 | 6 | ABL89850 | Abt189850 Human pol |
| 32 | 1855 | 42.1 | 1969 | 8 | ABT18375 | Abt18375 Aspergill |
| 33 | 1722 | 39.1 | 2546 | 2 | AAV44866 | Aav44866 Clone CTS |
| 34 | 1722 | 39.1 | 2546 | 5 | AAF98463 | Aaf98463 Human CDN |
| 35 | 1689 | 38.4 | 2510 | 5 | AAF93772 | Aaf93772 Human CDN |
| 36 | 1586 | 36.0 | 2537 | 4 | AAD08289 | Aad08289 Human sec |
| 37 | 1581 | 35.9 | 2547 | 4 | AAD08315 | Aad08315 Human sec |
| 38 | 1581 | 35.9 | 2660 | 6 | ABO54750 | Abq54750 Human ova |
| 39 | 1552.5 | 35.3 | 1114 | 4 | AAH99794 | Aah99794 Human pro |
| 40 | 1531 | 34.8 | 2284 | 4 | AAH18021 | Aah18021 Human CDN |
| 41 | 1370.5 | 31.1 | 1371 | 12 | ADP28508 | Adp28508 Human sec |
| 42 | 1253 | 28.5 | 787 | 4 | AAH07526 | Aah07526 Human CDN |
| 43 | 1200 | 27.3 | 764 | 5 | AAF93968 | Aaf93968 Primer sp |
| 44 | 907 | 20.6 | 507 | 3 | AAC01967 | Aac01967 Human sec |
| 45 | 872 | 19.8 | 2953 | 4 | ABL18224 | Abt18224 Drosophila |

ALIGNMENTS

RESULT 1

ADD94783
ID ADD94783 standard; cDNA; 2481 BP.

XX

AC ADD94783;

XX

DT 29-JAN-2004 (first entry)

DE Human SIM cDNA sequence.

XX

source of immunodominant MHC-associated peptide; SIMP; MHC;
major histocompatibility complex; human leukocyte antigen; HLA;
cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
lung cancer; intestine cancer; sarcoma; prostate cancer;
testicular cancer; breast cancer; melanomas; pancreatic cancer;
haematological cancer; immune response; lymphoid cell proliferation;
autoimmune disease; transplant rejection; SIMP-derived peptide; human;
gene; ss.

OS Homo sapiens.

XX

PN WO2003054008-A2.

XX

PD 03-JUL-2003.

XX

PF 18-DEC-2002; 2002WO-CA001967.

XX

PR 20-DEC-2001; 2001US-00028384.

XX

PA (COMP-) COMPATIGENE INC.

AFLOYLRDLTKQBFOTLFLGLVSLAAGAVFLSVLYLTGYIAPWSGRFYSFLMDTGY
AKIHPIIASVSEHQPTTWVSFFFDLHILVCTPPAGLWFCIKININDERVFGA"

ORIGIN

Alignment Scores: 5,94e-181 Length: 1664

Prod. No.: 2370.00 Matches: 443

Score: 99.77% Conservative: 1

Best Local Similarity: 99.77% Mismatches: 0

Query Match: 53.85% Indels: 0

DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x CQ782554 (1-1664)

QY 1 MetAlaGluProSerAlaArgGluSerIleHisGlySerSerLeuAsnSerSerProTyr 20
DB 28 ATGGCGAGCCCTCGGCCCGGAGAGACAGCAAGTCGTCCTCAACTCGTCCCCGTGG 87
QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
DB 88 AGTGGCCTCATGGCCCTGGGAAACAGCGCCGACACCGCCGCGCGGCCCGCCAGTGC 147
QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProCProLysProAlaProAlaGlyLeuSerGly 60
DB 148 GGGCACACAGCG 207
QY 61 GlyLeuSerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
DB 208 GGGCTGTGCGACCGCGCGTGGGTGGCAGTCGCTCTCTCCCTCCACCATCTCTCTCC 267
QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
DB 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAGCATCATCCAC 327
QY 101 GluPheAspProTyrPheAsnTyrArgSerThrHisIleAlaSerHisGlyPheTyr 120
DB 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTCATCTCATGGTTCAT 387
QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTyrTrpProLeuGlyArgIleValGlyGly 140
DB 388 GAATTTTAAATGGTTTGGTAAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGT 447
QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
DB 448 ACTGTTTACCCAGGTTGATGATACCGCTGCTTATTCATGGATTTTAAATACATGG 507
QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
DB 508 AACATAACTGTTTACATAAGAGACGTGTGTGTCTTCTGCACCACTTTTACCGCCTT 567
QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
DB 568 ACATCTATATCTACTTCTCTTACAAAGAGAACTTTGGAAACCAAGGAGGAGACTTTTA 627
QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
DB 628 GCTGCTTGTATTATGTCTATTGACAGGCTACATATCTGGTCAGTAGTGGATCCTTT 687
QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTrpLeuTrpValLysSer 240
DB 688 GATATAGAGGCGTATGCTATTTTTCACCTTCAGTTCACATCTATTATGGGTAATCT 747
QY 241 ValLysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
DB 748 GTAAAAACCTGGCTAGTTTGGACAATGTGCTGCTTATCTTATTTCTATATGCTC 807
QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
DB 808 TCTGCTTGGGCTGTTATGATTTATCATCTTATTCACCTGCATGATTTTGTGTTG 867
QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
DB 868 TTAATGATCAGAGATACAGCAAAAGAGTCTACATAGCATATAGCATTTCTACATTGTG 927

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
DB 928 GGTTTAATATATCAATGCAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAGTGAA 987
QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
DB 988 CACATGGCAGCTGCAGGTGCTTTGCAATGCTGCAGCTTATGCTTTCTTGCAGTATCTG 1047
QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
DB 1048 AGAGACCGGATTAACAAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1107
QY 361 AlaGlyValAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380
DB 1108 GCAGGTGCTGTGTTCTTGTAGTGTCTATCTATCTTACCTTATCAGGTTACATTGCACCATGG 1167
QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIle 400
DB 1168 AGTGGCAGGTTTTTATTCATTTGGGATACCTGGGTATGCAAAAATACACATTCCTCAATTATT 1227
QY 401 AlaserValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
DB 1228 GCATCAGTGTCTGAGCATCAACCTAGCACTTGGGTGTCTTTCTTCTTGTATCTACATATT 1287
QY 421 LeuValCysThrPheProIleGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
DB 1288 CTTGTATGTACCTTCCAGCAGGCTTTGGTTCCTGATCAAAAATATCAGCATGAAGA 1347
QY 441 ValPheValAla 444
DB 1348 GTATTTGGTGTCT 1359

Search completed: December 15, 2004, 05:31:26

Job time : 9979.01 secs

```

QY 344 uThrLeuGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaVa 364
DB 32246 CACCGGCAACAATCATGATCTCTCTTTGGGAGGATTCACCAATT--TTGGAGGCTCT 32302
QY 364 lPheLeuSerValIleTyr--LeuThrTyrThrGlyTyrIleAlaProTyrSerGlyAr 383
DB 32303 CTCGGTGTGTTTATTCGATCTGGTGGGAGATATGCTCCATCTCTGGAGC 32362
QY 383 gPheTyrSerLeuTyrPheThrGlyTyrAlaLysIleHisIleProIleIleAlaSerVa 403
DB 32363 TTCTACTCTCTGGGACACCGGATACCGGAGAAATTCATATTCCAATCATGCTCTCTGT 32422
QY 403 lSerGluHisGlnProThrTyrPheValSerPhePheAspLeuHisIleLeuValCy 423
DB 32423 CTCGAAACATCAGCAACACCTGGGTATCATTTCTTCGATCTCCATATTCACGCTGC 32482
QY 423 sPhePheProAlaGlyLeuTyrPheCysIleLysAsnIleAsnAspGluArgValPhe-- 442
DB 32483 TGTGTTCCAGTTCGGATTTGGTATTTGATTAAGAAGGTCAACGATGAGCGTGTTCAG 32542
QY 443 -----ValAlaLeuTyr 446
DB 32543 TAAGCTTAACAATCTGGATATGAATATTTTAAAAATTCGTTTTTCATGTTATCTGTA 32602
QY 446 lAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPr 466
DB 32603 CGCTGTGTCTCGGTTTACTTCGCGGAGTCATGTCCTCATGCTCACCTCACTCC 32662
QY 466 oValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAs 486
DB 32663 AGCTGTTGTGTCTCGCGGTATCGGATTTCTTACACTTCGAGAAATATCTTGAAGA 32722
QY 486 pAspMetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArgAs 506
DB 32723 CGAAGAAACAAAGAGCGTTCTTCTTCTCAGTCAGGAACCAACAAAGGAGGAGAG 32777
QY 506 nGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys----- 519
DB 32778 -----TTGTACCATATAAGCCGCCAAGAA-CGTCAAATCCAGGTTACCTGAAAGTG 32826
QY 519 ----- 519
DB 32827 TTAACAGCAGTTGTCAGCTGTCAGTGTGTTGTCCTTGTCTGTCCTGTTATGTGTTTC 32886
QY 519 ----- 519
DB 32887 TATAGCCTCATGTTTTTTTAAATTAATAAAGCGTTGTTTCCGTTGTTTATATTTAGTT 32946
QY 520 -----HisAlaThrGluGlnGluLysThrGluGluGl 530
DB 32947 TTATATGTGGAATTACTGTAATATGATTTTTCAGACGCCAAGCATGGAGATGAGTCCGG 33006
QY 530 yLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPhe 550
DB 33007 AGTTTCTTCTAATGTGGTACATCATTCATCATCTTGTGATCTTCTACTCATGTT 33066
QY 550 aAlaValHisCysThrTyrPheValThrSerAsnAlaTyrSerSerProSerValValLeuAl 570
DB 33067 GCTTGCTCATGCTACCTACGTACATCAACAGCGCTATTTCTCATCCATCTGCTGCTCCA 33126
QY 570 aserTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTyr 590
DB 33127 AAGTCTACTAACCAAGGAGACCGTATCATCATCATGAGGATTTCCGTGAGGCTTATCACTG 33186
QY 590 pLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTyrTyrAspTyrGlyTyrGl 610
DB 33187 GCTCGCGGAGAACACAGCTGATGATGCTCGCGTGTATGTCATGCTGGGATTTATGATATCA 33246
QY 610 nIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTyrAsnAsnSerHi 630
DB 33247 AATCCAGGAATGCGCAACCGTACTACATAGTTGACAAACAATACCTGGACAACTCGCA 33306
QY 630 sIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetAr 650

```

33307 CATTGGCCCTTGTGGAAAGACCATCTCTCAATGAATACAGCCCTTACGAAATATGAC 33366
 650 gThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAs 670
 33367 CGAACTCGAGGTTGACTACATCTCTTTATTTGGAGGAGTATTGGATATTCGAGA 33426
 670 pAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGluHisProLysAspIle 690
 33427 TGATATCAACAATCTCTTGTGATGTTCTGATTCGATTCACACAGGAGAACCCGAAAGATAT 33486
 690 eArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerPr 710
 33487 CCGCAGGAGAACTACTTCCATCGACTGGTGAATATATCCACTGGAGCGGTGCTTCTGA 33546
 710 oThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLe 730
 33547 AACTATGCTCAATTCGCTCATGTACAAATGAGCTATTATCGTTTGTGAACTCGTGT 33606
 730 uAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAs 750
 33607 TGGATACACACAGGCTGGTGGCTTCGATAGAACTCGTGAATATGTAATTCGAAAGAAAGA 33666
 750 pIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTyrLeuValArgIleTyr 770
 33667 TATCACTCTTGAATACATAGAAGACATATACACGAGAACTGGCTTGTGCTATCTA 33726
 770 rLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsnIle 790
 33727 CRAAGCTAAGAAGCTTCCAAACAGACCAACAGTGAAGAGTGAAGAGGTACCATCCCAAT 33786
 790 ePheProLysGln-----LysTyr-----Le 797
 33787 TAAGGGAGAGAGAGCCACACAGGTAAACGATTAAATAATATTCCTACTATTGTTAATAATTA 33846
 797 uSerLysLysThrThrLysArgLysArgGlyTyrIleLys 810
 33847 TAAACGATTGCAAGGTAAAAACAAACAAAGAGGATGATTCCA 33886

RESULT 15
 CQ782554
 LOCUS 1664 bp DNA linear PAT 17-MAR-2004
 DEFINITION Sequence 2694 from Patent EP1396543.
 ACCESSION CQ782554
 VERSION CQ782554.1 GI:45502338
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 TITLE Primers for synthesizing full length cDNA clones and their use
 JOURNAL Patent: EP 1396543-A 2694 10-MAR-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 Location/Qualifiers
 1..1664
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 28..1362
 /notes="unnamed protein product"
 /codon_start=1
 /protein_id="CAF85831.1"
 /db_xref="GI:45502339"
 /translation="MAEPSPAPESKHSLSNSPWSGLMALGNSRHGHGPGCAOAHKA
 AGGAAPPKAPAGLSGSLQSPAGQSLSTFLFLAWLAGSSELSFAVIRFESIIHEF
 DWFNYPRTSHLASHGFLVFLNWPDERAWYPLGRIVGVYVPLMTAGTTHWILNTL
 NITVHRDVCVFLPTAFSLTSTELLRELWNOAGLLAACATVPGVIGRSVAG
 SPENGLATFALQTYVLWKSIVTKTSVFTWMCCLSYFVWMSAWGVFIINLIPLH
 VFVLLMQEYSKRVIAVSTFYIVGLIILSMQIPFVGQPIRTSHVHAAGVFAALLQAY

```

/db_xref="wormBase:T12A2.16b"
/translation="MRKKWTIVLCILQATFSLYHDSERKSLMERPKSHKSCFSA
SFVILPWSQMHYPPKFTDRCPQSGIGLVCSSACFTLVERIDGEEYVSEQH
GVIRGCDRLFLGLDDVRNLSAYENQVRCHTRKLLRFLPLSGQDVTFFCSN
GDFNEHMLRELSNSFQHSVILLVATWILIF"
join(3241..3365,3586..3740,6392..6558,6795..6926)
/gene="hot-4"
/standard_name="T12A2.16a"
/notes="coded for by the following C. elegans cDNAs:
AP324056"
/codon_start=1
/product="Homolog of odr-2(two) protein 4, isoform a"
/protein_id="AAMI5600.1"
/db_xref="GI:20198878"
/db_xref="wormBase:T12A2.16a"
/translation="MRKKWTIVLCILQATFSLYHDSERKSLMERPKSHKSCFSA
SFVILPWSQMHYPPKFTDRCPQSGIGLVCSSACFTLVERIDGEEYVSEQH
GVIRGCDRLFLGLDDVRNLSAYENQVRCHTRKLLRFLPLSGQDVTFFCSN
GDFNEHMLRELSNSFQHSVILLVATWILIF"
complement(7287..12051)
/gene="T12A2.15"
/complement(join(7287..7364,7412..7520,7644..7735,
7844..7979,8026..8132,9511..9663,9726..9884,9932..10111,
10156..10458,10598..10687,10735..10911,10953..11156,
11304..11499,11832..11919,11982..12051))
/gene="T12A2.15a"
/standard_name="T12A2.15a"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
YK194a9.3, YK194a9.5, YK195a9.5, YK181e11.3, YK181e11.5"
/codon_start=1
/product="Hypothetical protein T12A2.15a"
/protein_id="AAC24452.2"
/db_xref="GI:20198877"
/db_xref="wormBase:T12A2.15a"
/translation="MSQSLVPLVGSAILSTFPFLGKWDYSFWVLLIIIASVTKS
FLYKRRERRIARATAREVIMQAQLDPAWQVPTDERVEWLNKIVHQLMYVYG
EYTFTRNDIIPQVAKQFMFKFTKMDMDI:PCRGGIKVYTNNGHDLIIVD
MDVAYGADFTVTCGCTGGMNIQFSGLKRALKPLPYPVWGVSGVSTPLEMPKM
DFNLTGEMVPELGLDARSINSQIAALCVNEIIVPLAPVDVTVLVFPBPDG
VRLIIEAKNLENDISPTKKGSDPAEIOVGSOFFKTRTIDDLNPNWEXPEAV
VDOAGKOLLELDFEDQGDSEGLSVDLKLVOAKGTIDKWYPLEGKGGDLHIKA
TWNLSLEHLEKQEAEGWQADFIHSAALLMYIDSVALPFPKSLFSPFVEV
SLGKETORTPVKNPFLPKSFLFVRHLEQELKFEAVDDGTRRSLSGLNIPLT
LLKEPNLEQOQHMLTVGVQSPFVITIRMLRSDSQSLNHRNSLRGLFRS
KHENKREVDENRGEIEQIDFDLNLQKIALIRCDLMTFKQOCNRYSVKVL
VALDGNKEVFKKTPAKTRHPHFNDHVEIDINPSDLNLNKHVINVKDDTNYGTFVA
KPLVGLCLEIRLDSLMNRQLSORWIPLSVERK"
complement(join(7287..7364,7412..7520,7644..7735,
7844..7979,8026..8132,8386..8532))
/gene="T12A2.15"
/standard_name="T12A2.15b"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
YK349e10.5"
/codon_start=1
/product="Hypothetical protein T12A2.15b"
/protein_id="AAC24451.1"
/db_xref="GI:3258591"
/db_xref="wormBase:T12A2.15b"
/translation="MNELSAPFIERANGKGVGVVPAKTVNQDAADEVEIKLN
VDNQMLRSPQSGLSHSGNSRLGRLFRSHEMKKRETRADNRGEIEQIDFDL
VNQKIALIRCDLMTFKQOCNRYSVKVLALDGNKEVFKKTPAKTRHPHFND
HVEIDINPSDLNLNKHVINVKDDTNYGTFVAKPLVGLCLEIRLDSLMNRQLSORWIPLS
VERK"
complement(16387..22005)
/gene="T12A2.5"
/complement(join(16387..16422,16786..17621,17892..18452,
18948..19052,19995..20008,20190..20562,20859..20951,
21039..22005))
/gene="T12A2.5"
/standard_name="T12A2.5"
/notes="contains similarity to Oryza sativa Putative myosin

```

heavy chain-like protein.; TR:Q94DC2; coded for by the following C. elegans cDNAs: CEES524F"

Alignment Scores:

| Pred. No.: | 1-02e-197 | Length: | 54118 |
|------------------------|-----------|---------------|-------|
| Score: | 2605.00 | Matches: | 502 |
| Percent Similarity: | 70.61% | Conservative: | 101 |
| Best Local Similarity: | 58.78% | Mismatches: | 139 |
| Query Match: | 59.19% | Indels: | 113 |
| DB: | 3 | Gaps: | 8 |

US-10-028-384-2 (1-826) x U13019 (1-54118)

| QY | 63 | SerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleLeuPheLeuAlaTrpLeu | 82 |
|--|-------|--|-------|
| DB | 31346 | AGTAGTCGGCTCGAGCGACACTCTTCTTACCATCGTCGACTGGCTCGCATGGTTC | 31405 |
| QY <td>83</td> <td>AlaGlyPheSerSerArgLeuPheAlaValIleAraPheGluSerIleIleHisGluPhe</td> <td>102</td> | 83 | AlaGlyPheSerSerArgLeuPheAlaValIleAraPheGluSerIleIleHisGluPhe | 102 |
| DB | 31406 | GTCGGATTGCGCTCCCGCTTTTCGCCATCGTCCGCTTTGAGTCGATTATCCACGAGTTC | 31465 |
| QY <td>103</td> <td>AspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe</td> <td>122</td> | 103 | AspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe | 122 |
| DB | 31466 | GATCCATGGTTAACTATATCGCTACACATCATATGTTCAACATGGTTCTATAAATTC | 31525 |
| QY <td>123</td> <td>LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal</td> <td>142</td> | 123 | LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal | 142 |
| DB | 31526 | TTGAAATGGTTGATGAGCGCTCGGTATCCACATTCGCTGCTGATGTCGGAGGAACTGA | 31585 |
| QY <td>143</td> <td>TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle</td> <td>162</td> | 143 | TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle | 162 |
| DB | 31586 | TATCCAGGACTCATGTGTAACATCTGGATTAACTCTGATTTTGATTTGATTCCTTAACATTC | 31645 |
| QY <td>163</td> <td>ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer</td> <td>182</td> | 163 | ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer | 182 |
| DB | 31646 | CATGTTACATTCGTAAGTTTCGGTATTTCTGGCTCCAAACATCTCTGGCTGACTGCC | 31705 |
| QY <td>183</td> <td>IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuLeuAla</td> <td>202</td> | 183 | IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeuLeuAla | 202 |
| DB | 31706 | ATTGCCACCTATCTACTTAAAGAGTTGTGGAGCCCGAGGAGCTGGTCTTTTTCGCCCT | 31765 |
| QY <td>203</td> <td>CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn</td> <td>222</td> | 203 | CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn | 222 |
| DB | 31766 | TGCTTATGCTATTTCTCCAGGATATACCTCCCTTCAGTTGCTGGATCTCATGATAAC | 31825 |
| QY <td>223</td> <td>GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSerValIys</td> <td>242</td> | 223 | GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSerValIys | 242 |
| DB | 31826 | GAGCGAATTGCCATCTTCTCTTCAATTCACGATATTTATTTATGCTGGAATCGCTCAAG | 31885 |
| QY <td>243</td> <td>ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerIle</td> <td>262</td> | 243 | ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerIle | 262 |
| DB | 31886 | ACCGGATCGGATATGTCGGCTTCGTTATGTCGCCCTTTTCACTTTTACATGTTTCGCCA | 31945 |
| QY <td>263</td> <td>TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu</td> <td>282</td> | 263 | TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeu | 282 |
| DB | 31946 | TGGGAGGATATGTTTTCATTAACATTGATTCACCTTCAGCTTCGCTTCATCATC | 32005 |
| QY <td>283</td> <td>Met-----Cln</td> <td>284</td> | 283 | Met-----Cln | 284 |
| DB | 32006 | ATGGGTAATTAATAAATAAATTAATGCAATTTATCAGTCGCAAAATTTTATTTCAGG | 32065 |
| QY <td>285</td> <td>-ArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLe</td> <td>304</td> | 285 | -ArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLe | 304 |
| DB | 32066 | TCGCTACTGTCGCCGCTCTGTTGTTTCTTACACATCGTCTTATTGCTTGCACCATCT | 32125 |
| QY <td>304</td> <td>uSerMetGlnIleProPheValGlyPheGlnPheIleArgThrSerGluHisMetAlaAl</td> <td>324</td> | 304 | uSerMetGlnIleProPheValGlyPheGlnPheIleArgThrSerGluHisMetAlaAl | 324 |
| DB | 32126 | CTCGATGCAAGTTCCTATTTGTAGGATTCACACCATCTGCTACTTCCGAGCACATCCGACG | 32185 |
| QY <td>324</td> <td>aAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLe</td> <td>344</td> | 324 | aAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLe | 344 |
| DB | 32186 | TTTGGAGTATTGGTCTTCTTCAATTTGCTCTCATGCATGCTAGCTAGCTCGCAACCGCAT | 32245 |

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, GenesScan gene prediction.

FEATURES

source
1..1734
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:358725"
/tissue_type="mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MNV."
/clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1734
/gene="1300006C19Rik"
/note="synonym: 3imp"
/db_xref="LocusID:68292"
/db_xref="WGI:1915542"
1..1549
/gene="1300006C19Rik"
/codon_start=2
/product="1300006C19Rik protein"
/protein_id="AAH03206.1"
/db_xref="GI:13096824"
/db_xref="LocusID:68292"
/db_xref="WGI:1915542"
/translation="GFQPIRTSEHMAAGVAFLLQAVFLVLDRLTKQBPQTLFFLGVSLAGAVFISVLYLYNTGYIAPWSGRFYSIMDTGVAKIHIPLIASVEHQPTTWSFFPDHLILVTFPAGLWFCIKINDEKRNPNLYDAGVAFAGVWVRLMLITFPVCMLSAIAISVFELGDDMKRNPPVEDSDSDKRNPNLYDAGVAFAGVWVRLMLITFPVCMLEPGLNKSIVITMLMLMFAVCTWTNSAYSPVLSYNHSDGTNILDPREA YFWLRQNTDEHARVMSWDYGYIAGMANRTLLVNDNTWNSHIALVGRKMSNETAA YKIMESLDVLYLVIFGVIGYSGDDINKFLMWRIAEGEHKPDIREGYFTQOGEFRVDKAGSPDLLMLVYKSYVFGMQLDFRTPGFDRFTRNAETGNKDIKFKLEKRAFTSEHLVRLVYKVPDNRETLGHKPRVNI VPKQYLSKTKTKRGRGVNKLAVFKKGK KTSKKTIV"
misc_feature
1..1249
/gene="1300006C19Rik"
/note="STT3; Region: Oligosaccharyl transferase STT3 subunit. This family consists of the oligosaccharyl transferase STT3 subunit and related proteins. The STT3 subunit is part of the oligosaccharyl transferase (OTase) complex of proteins and is required for its activity. OTase transfers a lipid-linked core-oligosaccharide to selected asparagine residues in the ER"
/db_xref="CDD:pfam04516"

ORIGIN

Alignment Scores:
Pred. No.: 1.03e-205 Length: 1734
Score: 2678.00 Matches: 505
Percent Similarity: 98.45% Conservative: 2
Best Local Similarity: 98.06% Mismatches: 8
Query Match: 60.85% Indels: 0
DB: 10 Gaps: 0
US-10-028-384-2 (1-826) x BC003206 (1-1734)
Qy 312 GlyPheGlnProileArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeu 331
Db 2 GGATTTACGCCAATCAGAACAGCGAGGCACATGCGAGTGTCTTTGGCTGCTG 61
Qy 332 GlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrlysgInGluPheGlnThr 351
Db 62 CAAGCTTACGCTTTTTCAGTATCTGAGACCGGTGACAAACAGGAGTTCACAGCC 121
Qy 352 LeuPheLeuGlyValSerLeuAlaAlaGlyValPheLeuSerValIleTyrLeu 371
Db 122 CTTTCTTTTGGGTGTCTACTAGTGCAGGCGCTGTGTCTTCTTAGTGTCATCTATCG 181
372 ThrTyrThrGlyTyrIleAlaProTrpSerGlyAArgPheTyrSerLeuTyrPheThrGly 391
Db 182 ACATACACAGGTATATATTCACCATGGAGTGGCAGGTTTTTATTTTACATATGGGATCTGGG 241
Qy 392 TyrAlaIleValIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrp 411
Db 242 TATGCAAAATACACATTCCTCAATATTGTCATCAGTGTCTGAACATCAGCCTACGACATGG 301
Qy 412 ValSerPhePhePheAspLeuHisIleLeuValCysThrPhePheAlaGlyLeuTrpPhe 431
Db 302 GTGTCTTCTCTTCTTCTATCATATATCTTGTATCTACCTTCCCGCAGGCGCTATGGTTC 361
Qy 432 CysIleIleValSerIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaVal 451
Db 362 TGCATCAAAATATCAACGATGAAGAGATTTTTCGCTCTGTATGCGATCAGTGTGCTGTC 421
Qy 452 TyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeu 471
Db 422 TACTTTGCGGAGTGTATGCTGGCGCTGTATCTGACTCTGACCCCGCTCTGTGATGCTG 481
Qy 472 SerAlaIleAlaPheSerHisValPheGluHisTyrLeuGlyAspMetLeuValArgGlu 491
Db 482 TCGGCCATCGCTTCTCCATGTTTTTGGAGCACTATTTGGGGGATGACATGAAAGGAA 541
Qy 492 AsnProProValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsnLeuTyr 511
Db 542 AACCCACCTGTGAGGACACGAGTGTATGAGGATGACAAAGAAACCCAGGAACTTGTAT 601
Qy 512 AspIleAlaGlyValArgIleValArgIleValArgIleValArgIleValArgIleVal 531
Db 602 GACAAAGCAGGTAAAGTGAAGAGCATGTGACAGAGCAGAGAGAACCTGAAGAGGGCTTG 661
Qy 532 GlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetLeuMetPheAla 551
Db 662 GGGCCCAACATCAAAAGCATGTGACCATGCTGTATGCTCATGCTCTGATGATGCTGCG 721
Qy 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSer 571
Db 722 GTCCACTGCACTGGGTGCACAGCAACGCTTACTCCAGTCCAGTGTGCTGCTTGGCTCC 781
Qy 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeu 591
Db 782 TAGCAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 841
Qy 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
Db 842 AGACAAAACACGGATGAACACGCGCGGTCTATGCTGTGGTGGACTACGGCTATCAGATT 901
Qy 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631
Db 902 GCTGGCATGGCCCAACAGGACCACTCTGGTGGTAAACAACACCTGGAAACACAGCCATC 961
Qy 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
Db 962 GCATGCTCGGAAACCTATGCTTCCATGAACAGCGCGCTATANAATCATGAGGTCC 1021
Qy 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 671
Db 1022 CTTGATGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1081
Qy 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691
Db 1082 ATCAACAAGTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1141
Qy 692 GluSerAspTyrPheThrProGlnGlyGluPheArgValAspValIleGlySerProThr 711
Db 1142 GAAGGTGATCTATTTTACCAGGAGGAGTTCAGAGTAGAACAAGCTGGGTGCTTCTTACT 1201
Qy 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
Db 1202 CTGTAAACCTGCCTTATGATATAAATGTACATACATACAGATTGGAGAAATCAGCTAGAT 1261
Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIle 751


```

Db 128698 TACACACAGAACACTGCTTGTTCGCACTATATAGGTGAAGAGCCGCGATGATGATCAAT 128699
Qy 778 ArgGluThrLeuAspHisLeuProArgValThrAsnLeuPheProLysGlnLysTyrLeu 797
Db 128638 AGACCATCACTGAAGACCAAGAGAGAACG-----ATTCTCCAGCAAACTTCATT 128588
Qy 798 SerLysLys-----800
Db 128587 TCGAGAAAGTAGTAGTACCTGCAGCTCCCAATGNACTTCAATTAACAACTGATCTTTG 128528
Qy 801 ----ThrThrLysArgLysArgGlyTyrLeuLysAsnLysLeuValPheLysLysGlyLys 819
Db 128527 CAGAACTCTAACCGCTGCAAGGCTACATACGAAACCGACGGTGTGTGTTAAGGAA 128468
Qy 820 LysLysSerLysLysThr 825
Db 128467 CGAACCTTGAAATAACC 128450

RESULT 12
AC018145/c
LOCUS
DEFINITION
AC018145
ACCESSION
VERSION
AC018145.1 GI:6553046
KEYWORDS
HTG: HTGS_PHASE2.
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 75650)
AUTHORS
Adams,M. and Venter,J.C.
TITLE
Direct Submission
JOURNAL
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10213542 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..75650
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Alignment Scores:
Pred. No.: 1,61e-212 Length: 75650
Score: 2791.50 Matches: 556
Percent Similarity: 76.12% Conservative: 88
Best Local Similarity: 65.72% Mismatches: 115
Query Match: 63.43% Indels: 91
DB: 2 Gaps: 10

US-10-028-384-2 (1-826) x AC018145 (1-75650)
Qy 63 SerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrLeuPheLeuAlaTyrLeu 82
Db 21460 AGCAAGTGGCTGGCTACAGACGCTATACCTTCGCCATCTCTTAATCGGCTGGCTG 21401
Qy 83 AlaGlyPheSerSerArgLeuPheAlaValLeuArgPheGluSerLeuLeuHisGluPhe 102
Db 21400 GCCGGAATTTCTCTCGCCTCTTCGCCGCTATCGCTTCGAGTCGATTCATTCAGTGT 21341
Qy 103 AspProTyrPheAsnTyrArgSerThrHisLysLeuAlaSerHisGlyPheTyrGluPhe 122
Db 21340 GATCCGTGGTTCATCACTACCGGCCACCCCTACATGGTGGAGATGTTGGTACAACTTC 21281
Qy 123 LeuAsnTyrPheAspGluArgAlaTyrTyrProLeuGlyArgLysLeuValGlyThrVal 142

```


| | | | |
|----|--------|---|--------|
| QY | 63 | SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpIleu | 82 |
| Db | 171786 | AGCAAGTGGGTGGTTCAGCAGAGCGTAAACACCTTCGCCATCTGCTAAATCGCCTGGCTG | 171727 |
| QY | 83 | AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe | 102 |
| Db | 171726 | GCGGATTTTCCTCTCGGCTCTTCGCGTCATCGGTTCAGTCGATTATCCATGAGTTT | 171667 |
| QY | 103 | AspProTrpPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrGluPhe | 122 |
| Db | 171666 | GATCCGTGGTTCACATACCGGCGCACCGCTACATGTCGAGAATGGTGGTACAACTTC | 171607 |
| QY | 123 | LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrVal | 142 |
| Db | 171606 | CTCAACTGGTTCGACGAGCGGCATGTTATCCGCTCGCAGGATTGTGGCGGTACCGTC | 171547 |
| QY | 143 | TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle | 162 |
| Db | 171546 | TATCCGCGCTGATGATTACGTCCGGCGGAATCAATTGGCTGTGCACGTACTCAACATA | 171487 |
| QY | 163 | ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer | 182 |
| Db | 171486 | CCGGTCCATATTCTGATCATTCGGTGTCCTCGCGCGCATCTCACTGGCGCTGACCTCC | 171427 |
| QY | 183 | IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAla | 202 |
| Db | 171426 | ATCTCCACCTACTCTCTCAACAAGAGCTGTGGTCCGCGCGCGCGCGCTCTTCGCGCGC | 171367 |

Db 3636 CTCAACTGTTTCGACGCGCGCATGGTATCCCGTGGCAGGATGTTGGCGCGTACCGGTC 3577
 Qy 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIle 162
 Db 3576 TATCCCGGCTGATGATACGTCGCGCGGATCCATGGCTGCTGCACGTACTCAACATA 3517
 Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 Db 3516 CGGTCCTATATTCGATCATCTGGTGTCTCGCGCGCATCTTCAGTGGCTGACCTCC 3457
 Qy 163 IleSerThrPheLeuLeuThrArgGlyLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
 Db 3456 ATCTCCACCTACCTGCTGACCAAGAGCTGTGGTCCGCGCGCGCTCTTCGCGCGC 3397
 Qy 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 Db 3396 AGCTTCATCGCATCGTGGCTGGGTATCATCAGTAGTGGTGGCTGGATCGTACGATAAC 3337
 Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleValSerValIys 242
 Db 3336 GAGGCGATTCGATATTCGCTCGCATTCACCTACTCTCTGGTGGCTCAGTGAG 3277
 Qy 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 Db 3276 ACTGGATCGGTGTTCTGGTCCGCGCGCGCATCTTGGCTACTTCTACATGGTGGCGCC 3217
 Qy 263 TrpGlyGlyTyrValPheIleAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
 Db 3216 TGGGGTGGCTACGTGTTTCATCATCACTGATACCTCCCTGACCTCTCTGCTACTGCTCATT 3157
 Qy 283 MetGlnArgTyrSerIysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
 Db 3156 ATGGCAGGATCTCGCGCGCTCTGCTGACCGAGTACAGCACTTCACATCTCGGAGCTG 3097
 Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 Db 3096 CTGTTCTCATGCGATCCCTCTGTTGGATTCACACCGATACGACAGTGAACACATG 3037
 Qy 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 Db 3036 GCTGGCTGGGAGTGTGCTCTCTATGGCGGTGGCCACCTTCGCGCATTTGCGAGTCC 2977
 Qy 343 ArgLeuThrIysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGly 362
 Db 2976 GTGCTGTGCGCAAGAGTTCGGAAGCTGTTCATGTCGCGCGATGCTGGTGGCGTT 2917
 Qy 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProThrPheGly 382
 Db 2916 GCGCTGTTTGTGGCGGTGCTGCTCACCATGCTGGGCGGTGTCGCGCGCTGGAGTGA 2857
 Qy 383 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaIysIleHisIleProIleIleAlaSer 402
 Db 2856 CGCTTCTACTGCTGGGATCTGCTACGCGCAAGATCCACATTCCTCATCTGATCC 2797
 Qy 403 ValSerGluHisGlnProThrThrTrpValSerPhePhePheAspIleHisIleVal 422
 Db 2796 GTGCGGAGATCAGCCACCATGTTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2737
 Qy 423 CysThrPheProAlaGlyLeuTrpPheCysIleIysAsnIleAsnAspGluArgValPhe 442
 Db 2736 TGGCGCTTCCAGTGGGAGTGTGGTACTGATCATCAGCAGATCAACGCGAGCGGTTTC 2677
 Qy 443 ValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462
 Db 2676 GTGGTGTGTACGCGCATCAGTGGCGGTTTACTTCGCTGTGTGTGTGTGTGTGTGTGTG 2617
 Qy 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
 Db 2616 ACCCTCAGCGCGGTGGTGTGATCTGCGCGGAGTGGCTTTTTCGGGACTGTGTGATGTG 2557
 Qy 483 TyrLeuGlyAspAspMetLysArg-----GluAsn 492
 Db 2556 TTCCTGCAAGAGGATTCGTCTAAGCGAATGGGACACGCCATAGACGCCACCGAAGTG 2497

Qy 493 ProProValGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeu----- 510
 Db 2496 GATGAAGCTGAGGATTCATTGA-GAAGAAGACGCTGTACGACAGGTGAGTCTTCTACTA 2438
 Qy 511 -----TyrAspLysAlaGlyLysValArg 518
 Db 2437 ACACATCATCGGATTTGATTTAAATATACATCGCATTCGAGCTGGCAAGCTG--- 2381
 Qy 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
 Db 2380 AAGCATCGTACTAAGCATGATGCCAGCAGGATGACTGGCGTCAGCTCCAACTCAAGAGT 2321
 Qy 538 IleValThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpVal 557
 Db 2320 ATTGTTATTGTCCTTCTAATGCTGTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 2261
 Qy 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
 Db 2260 ACCAGCAATGCTTACTTCCAGTCCCTTCTGCTTGGCTTTCACACACAGTCAAGATGG 2201
 Qy 577 -----Thr 577
 Db 2200 TAGCACCAGTAGCATCATGCTCTTCTTAAACCTATCTAACCTTTTCTCCACTAGATCC 2141
 Qy 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGlu 597
 Db 2140 CGCAACATTTAGACGATTCAGAGAGGCTTACTACTGGCTTTCGACAGACCTGCGCAT 2081
 Qy 598 HisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 617
 Db 2080 GATGCTCGGCTGATGCTTGGTGGGATACGATACAGATAGCGGATGGGATGGCAACAGA 2021
 Qy 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
 Db 2020 ACCAGCTAGTGGATAATAATACGTGGAAACATAGTACATAGCGCTGGTGGCAAGGCA 1961
 Qy 638 MetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrVal 657
 Db 1960 ATGCTTTCACCGAGGAGAGTCTACGAAATATGATGATCTCTTGACGTGGATGATGTT 1901
 Qy 658 LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 677
 Db 1900 TTGGTATCTTGGCGGTGTGATCGCTATTCTGCGGATGATATCAACAGTCTCTGTGG 1841
 Qy 678 MetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThr 697
 Db 1840 ATGCTCGAATGCTGAGGAGAGCATCCCAAGACATTAAGGAAGGATTTACTTTACC 1781
 Qy 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
 Db 1780 GACCGCGGTGAATTCAGGCTAGATCCGAAGGTGCTCGGCGCTCTCACTGCGCTTATG 1721
 Qy 718 TyrLysMetSerTyrTrpArgPheGlyGluMetGlnLeuAspPheArgThr----- 734
 Db 1720 TACAAATTAAGCTACTACAGATTCGCGGAATTTGAAGTTGGACTACAGTA-AGCGGAAC 1662
 Qy 735 -----ProProGlyPheAsp 739
 Db 1661 ATTTCTCTCAGGTAGCGATGCACATACTGCTCTTTCAGAGGTGCTTCTGGATATGAT 1602
 Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAla 759
 Db 1601 CGCACAGTAACCGCTCTATTGGGAATTAAGACTTCGATCTGACCTACTCGAGGAGGCC 1542
 Qy 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAsp-----Asn 777
 Db 1541 TACACACAGAACACCTGCTTGTGCTGCTATAGGTTGAAGTGAAGAGCCGATGAGTTCAAT 1482
 Qy 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeu 797
 Db 1481 AGACCATCTCTGACAGCAGGAGGAGACG-----ATTCTTCCAGCAAACTTCATT 1431


```

Qy 577 ThrArgAenIleLeuAapAapPheArgGluAlaTyrPheTrpLeuArgGlnAenThrAap 596
Db 1655 TCCGCGACATTTTAGACGATTTTCAGAGAGCGCTTACTACTGGGTTTCGAGAACACCTGCC 1714

Qy 597 GluHisAlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAen 616
Db 1715 GATGATGCTCCGGTATGCTTGTGGATACCGATACCGATACCGGGAATGGCAAC 1774

Qy 617 ArgThrThrLeuValAapAenAenThrTrpAenAenSerHisIleAlaIleValGlyLys 636
Db 1775 AGAACGACGCTAGTGGATTAATAACCTGGAAACAATAGTCACATACGCTGCTGGTGGCAAG 1834

Qy 637 AlaMetSerSerAenGluThrAlaAlaTyrLysIleMetArgThrTrpLeuAapValAapTyr 656
Db 1835 GCATGCTTCAACCGAGAGAACTCTCAGAAATATGACATCTCTGACGTGGACATAC 1894

Qy 657 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAenLysPheLeu 676
Db 1895 GTTTTGGTGATCTTTGGCGGTGTCATCGCTATTCTGGCGATGATATCAACAAGTTCTGTG 1954

Qy 677 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
Db 1955 TGGATGCTCCGAATGCTGGAGGAGACATCCCAAGACATTAAAGGAAGCGATTACTTT 2014

Qy 697 ThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAenCysLeu 716
Db 2015 ACCGACCGCGGTGAATTCAGGGTAGATGCCGAAGGTGCTCCGCGCTGTCTCAACTGCCCT 2074

Qy 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAapPheArgThrProPro 736
Db 2075 ATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTACAGAGGTCCATCT 2134

Qy 737 GlyPheAapArgThrArgAenAlaGluIleGlyAenLysAspIleLysPheLysHisLeu 756
Db 2135 GGATATGATCGCACACGCTAACCGCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG 2194

Qy 757 GluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAap 776
Db 2195 GAGGAGGCTACACACAGACATCTGCTGTTCGCATCTATAGGTTGAGAGCGGCAT 2254

Qy 777 -----AsnArgGluThrLeuAenHisLysProArgValThrAenIlePheProLysGln 794
Db 2255 GAGTTCAATAGACCATCACTCAAGACCAAGAGAGAACG-----ATTCTCCAGCA 2305

Qy 795 LysTyrLysSerLysLysThrThrLysArgLysArgGlyTyrIleLysAenLysLeuVal 814
Db 2306 AACTTCATTTGAGNAAGAACTCTAAGGTCGCAAGGGCTACATAGAAACCGCGGT 2365

Qy 815 PheLysLysGlyLysLysLysSerLysLysThr 825
Db 2366 GTTGTAAAGGAAACGAACCTTGAAATAAAC 2398

RESULT 7
LOCUS CQ589353
DEFINITION Sequence 17111 from Patent WO0171042.
ACCESSION CQ589353
VERSION CQ589353.1 GI:41648215
KEYWORDS
SOURCE
ORGANISM Drosophila sp.
Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1
REFERENCE
AUTHORS Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
TITLE Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 0171042-A 17111 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
1..2699
/organism="Drosophila sp."
source

```

```

/mcl_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN
Alignment Scores:
Pred. No.: 2,5e-227 Length: 2699
Score: 2950.00 Matches: 557
Percent Similarity: 83.79% Conservative: 89
Best Local Similarity: 72.24% Mismatches: 113
Query Match: 67.03% Indels: 12
DB: 6 Gaps: 7

US-10-028-384-2 (1-826) x CQ589353 (1-2699)
Qy 63 SerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaTrpLeu 82
Db 107 AGCAAGTGGCTGGCTACAGAGCGCTATACCTTCGCATCTCTGCTAAATCGCTGGCTG 166

Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
Db 167 GCGGATTTTCTCTCGCTCTTCGCGCTCATCCGTTTCGAGTCGATTATCCATGAGTTT 226

Qy 103 AspProTrpPheAenTyrArgSerThrHisIleLeuAlaSerHisGlyPheTyrGluPhe 122
Db 227 GATCCGTGGTTCACTACCGGCGCACCGCTACATGGTGCAGATGGTGTGTACATTC 286

Qy 123 LeuAenTrpPheAapGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal 142
Db 287 CTCAACTGGTTCGACGAGCGCATGGTATCCGCTCGGCGAGATTGTGGCGGTACCGTC 346

Qy 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAenThrLeuAsnIle 162
Db 347 TATCCCGGCTGATGATACGTCGCGGCAATCCATTCGCTGCTGCAGCTACTCAACATA 406

Qy 163 ThrValHisIleArgAapValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
Db 407 CCGGTCCATATTCGTACATCTCGCTGTTCCTCGGCGCGATCTTCAGTGGCTGACCTCC 466

Qy 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuAlaAla 202
Db 467 ATTCCACCTACTCTGCTGACCAAGAGCTGTGTCGCGGCGCGCGCTCTTCGCGGCC 526

Qy 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAapAen 222
Db 527 AGCTTCATCGCCATCTGCTGGCTACATCAGTAGGTCGCTGGCTGAGTCTGACGATAAC 586

Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLys 242
Db 587 GAGGCAATGCCATATTCGCTTCAGTTTCACTACTCTCTGTTGGTGGCTCAGTGAAG 646

Qy 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
Db 647 ACTGGATCGGTGTTCTGGTCGCGCGCGCGCTTTTGTCTACTTCTACATGGTGTCCGCC 706

Qy 263 TrpGlyGlyTyrValPheIleIleAenLeuIleProLeuHisValPheValLeuLeuLeu 282
Db 707 TGGGTGGCTACGCTGTTCAATCACTGATACCCCTGCAGCTCTTGGTACTGCTCAT 766

Qy 283 MetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
Db 767 ATGGCAGAGTACTCGCGCGGTCTGTGTGACCACTACAGACCTTCTACATCTCTGGACTG 826

Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
Db 827 CTGTTCTCCATGCAATCCCTTCGTTGGGATTCACACCGATACGACACGAGTGAACATG 886

Qy 323 AlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlyTyrIleuArgAap 342
Db 887 GCTGCGTGGAGTGTGTTGTCTCTTATGGCCGCTGGCGCACCTTCGCGCATTTGACATCC 946

Qy 343 ArgLeuThrLysGlnGluPheGlnThrLeuPheLeuGlyValSerLeuAlaAlaGly 362
Db 947 GTGCTGTGGCGCAACGAGTTCCGGAAGCTGTTTCATCGTCGGCGGATTCTGTGGTGGCTT 1006

```

Db 2306 AACCTTCATTTCGAGAAAGAACTCTAAGCGTCGCAAGGCGCTACATACGAAACCGACCGGTT 2365

Qy 815 PheLysLysGlyValSerLysThr 825

Db 2366 GTTGTTAAGGAAGAACCACTTGAAATTAACC 2398

RESULT 6

AX799088

LOCUS AX799088 2417 bp DNA linear PAT 08-OCT-2003

DEFINITION Sequence 7 from Patent WO03054008.

ACCESSION AX799088

VERSION AX799088.1 GI:37605061

KEYWORDS Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1

REFERENCE Perreault, C. and McBride, K.

AUTHORS Mammalian SIMP protein, gene sequence and uses thereof in cancer

TITLE therapy

JOURNAL Patent: WO 03054008-A 7 03-JUL-2003;

COMPATIGENE Inc. (CA)

FEATURES

Source Location/Qualifiers

1..2417

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/note="STT3, Acc# AF132552"

ORIGIN

Alignment Scores:

Pred. No.: 2,148-227 Length: 2417

Score: 2950.00 Matches: 557

Percent Similarity: 83.79% Conservative: 89

Best Local Similarity: 72.24% Mismatches: 113

Query Match: 67.03% Indels: 12

DB: 6 Gaps: 7

US-10-028-384-2 (1-826) X AX799088 (1-2417)

Qy 63 SerGlnProAlaGlyTrpGlnSerLeuSerPheThrLeuPheLeuAlaTrpLeu 82

Db 98 AGCAAGTGGTGGTACAGAGCCATAACCTTCGCCATCTTCGCTGCTGCTGCTGCTG 157

Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102

Db 158 GCGGATTTCTCTCGCTCTTCGCGTCTTCGCGTCTTCGAGTCGATTATCCATGAGTTT 217

Qy 103 AppProTrpPheAsnTrpArgSerThrHisIleuAlaSerHisGlyPheTrpGluPhe 122

Db 218 GATCCGTGGTTCAACTACCGGCCACCGCTACATGTGTGAGAAATGGTTGGTACAACTTC 277

Qy 123 LeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyThrVal 142

Db 278 CTCACCTGTTTCGACGAGCGCATGTTATCCGTCGCGAGGATTTGGCGGCTACCGTC 337

Qy 143 TrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleuAsnThrLeuAsnIle 162

Db 338 TATCCCGGCTGATGATACGTCGCGGAAATTCATTTGGCTGCTGCGACGTACCAACATA 397

Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182

Db 398 CCGGTCCATATTCGTGACATCTGCGTGTTCCTGGCGCGATCTTTCAGTGGCTGACCTCC 457

Qy 183 IleSerThrPheLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAala 202

Db 458 ATCTCCACCTACTGCTGACCAAGAGCTGTGGTCCCGGGCGCGCGCTCTTCGCGGCC 517

Qy 203 CysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerPheAspAsn 222

Db 518 AGCTTCATCGCCATCGTGGCTGGTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGT 577

Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeuTrpValSerValys 242

Db 578 GAGGCAATGGCAATATTCGCTCGCTGAGTTCACTACTTCTCTGGTGGTGGTGGTGGTGGT 637

Qy 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetValSerAla 262

Db 638 ACTGATATTCGCTGCTGGTGGCGGCGGCTTTGCTACTTCTACTTCTACTTCTACTTCT 697

Qy 263 TrpGlyGlyTrpValPheIleIleLeuLeuIleProLeuHisValPheValLeuLeu 282

Db 698 TGGGTGGCTACGCTGTTTCATCAACCTGATACCTCTGCGACGCTCTTCTGCTGCTGCT 757

Qy 283 MetGlnArgTrpSerLysArgValTrpIleAlaTrpSerThrPheTrpIleValGlyLeu 302

Db 758 ATGGCAGGTACTTCGCGGCTGCTGCTGACGCTACGACCTTCTACTTCTACTTCTACTTCT 817

Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322

Db 818 CTGTTCTCCATGAGATCCCTTCGTTGGGATTCACCAACGATACGACGACGACGACGAC 877

Qy 323 AlaAlaAlaGlyValPheAlaLeuGlnAlaTrpAlaPheLeuGlnTrpLeuArgAsp 342

Db 878 GCTGGCTGGAGTGTTCGCTTATGCGCGTGGCCACCTTCGCGCAATTCGAGTCC 937

Qy 343 ArgLeuThrLysGlnPheGlnThrLeuPhePheLeuGlyValSerIleAlaAlaGly 362

Db 938 GTGCTGTCGCGCAACGAGTTCGGAAGCTGTTTCATCGTCGCGGAGTTCGTCGCGGCT 997

Qy 363 AlaValPheLeuSerValIleTrpLeuThrTrpThrGlyTrpIleAlaProTrpSerGly 382

Db 998 GCGCTCTTGTGGCGCTGCGTGGTCTCACCACCTGGCGTGGTGGTGGTGGTGGTGGT 1057

Qy 383 ArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIleAlaSer 402

Db 1058 CGCTTCTACCTGCTGGGATCTGCTACGACGACAGATCCACATTCCTCCATTCGATCC 1117

Qy 403 ValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIleLeuVal 422

Db 1118 GTGTCGAGCATCAGCCACCACTTGGTCTCTGTTCTCTTCTCTCTCTCTCTCTCTCTCT 1177

Qy 423 CysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPhe 442

Db 1178 TGGCGCTTCCATGGAGGTGGTACTGCATCAGCAGATCAACGACGAGCGGCTTTC 1237

Qy 443 ValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArgLeuMetLeu 462

Db 1238 GTGCTGCTGTACGCCATCAGTGGGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1297

Qy 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482

Db 1298 ACCCTACCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1357

Qy 483 TyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSerSerAspGlu 501

Db 1358 TTCCTGCAAGAGGATTCGTCTAAGCAATGGGACAGACCATTAAGCGGACGACCAAGATG 1417

Qy 502 Asp-----AspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArg 518

Db 1418 GATGAACCTGAGGATTCATTTGAGAGAGACGCTGTACGACAGGCTGGCAGCTG--- 1474

Qy 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537

Db 1475 AAGCATCGTACTAAGCATGATGCCAGCAGGATCTGGCGTCAGCTCCAACTCCAACTGAGT 1534

Qy 538 IleValThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpVal 557

Db 1535 ATTGTTATTTGGCGCTTCAATGCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTG 1594

Qy 558 ThrSerAsnAlaTrpSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576

Db 1595 ACCAGCAATGCTACTCAGTCCCTCCATTTGCTTGGTCTTCCACACAGTCAAGATGGA 1654

Qy 83 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 102
Db 158 GCGGAAATTTCTCTCGCCCTCTTCGCGCTCATCGGTTCGAGTCGATATCCATGAGTTT 217
Qy 103 AspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPhe 122
Db 218 GATCCGTGGTTCACACTACCGGCGCACCGCTACATGGTGCAGATGGTTCGTAACAATTC 277
Qy 123 LeuAsnTyrPheAspGluArgAlaTyrProLeuGlyArgIleValGlyIleThrVal 142
Db 278 CTCAACTGGTTCGACGAGCGGCATGGTATCGCTCGGACAGGATGGCGCGTACCGTC 337
Qy 143 TyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTyrPheLeuAsnThrLeuAsnIle 162
Db 338 TATCCCGGCTGATGATTAGTTCGCGCGAATCCATTGGCTGTGCACGTACTCAACATA 397
Qy 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
Db 398 CGGTCCTATTCGTGACATCTGCGTGTCTCGCGCGCATCTTCAGTGGCCCTGACCTCC 457
Qy 183 IleSerThrPheLeuLeuThrArgGluLeuTyrPheGlnGlyAlaGlyLeuLeuAlaAla 202
Db 458 ATCTCCACCTACTGCTGACCAAGGAGCTGTGGTCCGCGCGCGCTCTTCGCGGCC 517
Qy 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
Db 518 AGCTTCATCGCCATCGTGGCTGGTACATCATGATGCTGGTGGTGGTGGTGGTGGTGGT 577
Qy 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValIleValSerValIys 242
Db 578 GAGGCGCATGCTATTCGCTGCGCTGACCTACTCTCTGTTGGTGGCTCAGTGAAG 637
Qy 243 ThrGlySerValPheThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
Db 638 ACTGATCGCTGTTCTGGTGGCGCCAGCCGCTTGTCTACTTCTACATGGTGGTGGTGGT 697
Qy 263 TrpGlyGlyTyrValPheIleLeuAsnLeuIleProLeuHisValPheValLeuLeuLeu 282
Db 698 TGGGGTGGCTACTGTTTCATCAACCTGATACCCCTGCGACGCTCTGCTACTGCTCATT 757
Qy 283 MetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeu 302
Db 758 ATGGGAGGTACTCCCGGCTGCTGACAGCTGACAGCTTCTACATCTCTGAGGACTG 817
Qy 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
Db 818 CTGTTCTCCATGCAGATCCCTTCTGTGGATTCACACCGATAGCCACGAGTGAACATG 877
Qy 323 AlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
Db 878 GTTGGCTGGAGTGTGCTCTCTATGCGCGCTGGCCACCTTGGCGCATTTGGAGTCC 937
Qy 343 ArgLeuThrLysGlnGluPheGlnThrPhePheLeuGlyValSerLeuAlaAlaGly 362
Db 938 GTGCTGTCCGCAACGAGTTCGGAAGCTGTTCATCTGCGCGCATTTGCTGGCGCTT 997
Qy 363 AlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProThrSerGly 382
Db 998 GCGCTCTTGTGGCGCTGCTGGTCTACCATGCTGGCGCTTGTGGCGCGCTGGAGTGA 1057
Qy 383 ArgPheTyrSerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleLeuAlaSer 402
Db 1058 CGCTTCTACTGCTGTGGGATCTGGGTACGCCAAGATCCACATTCCTCCATTTGTCATCC 1117
Qy 403 ValSerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIleLeuVal 422
Db 1118 GTGTGGAGCATCAGCCACCATTTGTTCTGTTCTTCTGATCTGACATCTCTGGTG 1177
Qy 423 CysThrPheProAlaGlyLeuTyrPheCysIleLysAsnIleAsnAspGluArgValPhe 442
Db 1178 TCGGCTTCTCCAGTGGGAGTGTGATGTCATCAAGCAGATCAACACGAGCGGCTTTTC 1237
Qy 443 ValAlaLeuTyrTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 462

Db 1238 GTGGTGTGTACCCATCATGTCGGTTACTTCGCTGGTGTGATGGTGGTTCGATGTTG 1297
Qy 463 ThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHis 482
Db 1298 ACCCTCAGCGCGGTGTGTGATGTCGTGGCGGAGTGGCTTTTCGGGACTGTTGGATGTG 1357
Qy 483 TyrLeuGlyAspAsp---MetLysArgGluAsnProProValGluAspSerSerAspGlu 501
Db 1358 TTCTGCAAGAGGATTCGTCTAAGCGAATGGCGACAGCCATTAAGCGCACCGCAAGATG 1417
Qy 502 Asp-----AspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArg 518
Db 1418 GATGAAGCTGAGGATTCATTCAGAGAAGAGCGCTGTACGACAAGGCTGGCAAGCTG-- 1474
Qy 519 LysHisAlaThrGluGlnGluLysThrGluGlu---GlyLeuGlyProAsnIleLysSer 537
Db 1475 ARGATCGTACTAAGCATGATGCCAGCAGGATATCTGGCGTCACTCCAACTGAAAGAT 1534
Qy 538 IleValThrMetLeuMetLeuMetMetPheAlaValHisCysThrTyrVal 557
Db 1535 ATTGTTATTGCGCGCTTCTAATGCTGTGTGATGTTTCGCTGTCCACTGCACGTGGGTG 1594
Qy 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn---HisAspGly 576
Db 1595 ACCAGCAATGCGCTACTCCAGTCCCTCCATGCTGTGGCTTCCCAACAAGATGGA 1654
Qy 577 ThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTyrLeuArgGlnAsnThrAsp 596
Db 1655 TCCCGCAACATTTAGACGATTTACAGAGGCTTACTACTGGCTTTCGAGAACACTGCC 1714
Qy 597 GluHisAlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 616
Db 1715 GATGATGCTCGCGTATGCTTGGTGGGATACGATACGATAGGAGGAGGAGGAGGAG 1774
Qy 617 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 636
Db 1775 AGAACGACGCTAGTGGATAATAATACGTGGAAACATAGTACATAGCGCTGGTGGCAAG 1834
Qy 637 AlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyr 656
Db 1835 GCAATGCTTTCACCGAGGAGAGGCTTACGAAATATGACATCTCTGACGTGGACTAC 1894
Qy 657 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 676
Db 1895 GTTTTGGTGTCTTGGCGGTGTGATCGCTATTCTGCGCATGATATCAACAAGATTCCTG 1954
Qy 677 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPhe 696
Db 1955 TGGATGTCGGAATTCGTGGGAGAGCATCCCAAGGACATTAAGAAAGCGATTTACTTT 2014
Qy 697 ThrProGlnGlyPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeu 716
Db 2015 ACCGACCGCGTGAATTCAGGGTAGATCCGAAGGTGCTCGCGCTGCTCAACTGCTT 2074
Qy 717 MetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProPro 736
Db 2075 ATGTACAAATTAAGTACTACAGATTCGGGANTTGAAGTGGACTACAGGCTCATCT 2134
Qy 737 GlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeu 756
Db 2135 GGATATGATCGCACACGTAACCGCGCTCATTTGGGAATTAAGGACTTCGATCTGACCTAC 2194
Qy 757 GluGluAlaPheThrSerGluHisIleTyrLeuValArgIleTyrLysValLysAlaProAsp 776
Db 2195 GAGGAGGCTACACACAGACACTGGCTTCTGCTATATAGGTTGAAGAGCCGAT 2254
Qy 777 -----AsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGln 794
Db 2255 GAGTTCATATAGACCATCTACTGAAGCAAGAGAGAGACG-----ATTCTCCAGCA 2305
Qy 795 LysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuVal 814

```

Db 1805 ATATTAGATGATTTAGAGAGCGCTACTTTGGCTGAGACAAACACGATCAACAGCC 1864
Qy 600 ArgValMetSerTTPtTpAspTyrGlyTyrGlnLeuAlaGlyMetAlaSerThrThr 619
Db 1865 CGGGTCATGTCGTGGTGGACTAGCTGATCAGATTGCTGGCATGGCCACAGGACCAT 1924
Qy 620 LeuValAspAsnAsnThrTTPtTpAsnSerHisLeuAlaLeuValGlyLysAlaMetSer 639
Db 1925 CTGGTGGATAACAACACCTGGAACAACAGCCACATGCTGCTGCTGCGAAAGCTATGCT 1984
Qy 640 SerAsnGluThrAlaAlaTyrLysLeuMetArgThrLeuAspValAspTyrValLeuVal 659
Db 1985 TCCATGAAAGCGCGGCTATAAATCATGAGTCCCTGATGCTGATGATGCTGCTGCT 2044
Qy 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetVal 679
Db 2045 ATTTTCGAGAGATGATGCTGCTATTCGCGGAGCATATCAACAAGTTCCTCTGGATGCT 2104
Qy 680 ArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
Db 2105 AGGATAGCTGAAGGAGCATCCCAAGACATCCGGAAGTCACTATTTCACCCAGCAG 2164
Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnLysLeuMetTyrLys 719
Db 2165 GGAGAGTTCGAGTAGACAAGCTGGGTCTCTACTCTGTAAACTGCCTTATGTATATAA 2224
Qy 720 MetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAsp 739
Db 2225 ATGTCATACTACAGATTGGAGAAATGACAGTAGATTTCGCATCCCCAGGCTTTCAC 2284
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759
Db 2285 CGAACACATATGCTGAGATTGGAATTAAGACATTAATTCAGCATTTGGAGGAAGCT 2344
Qy 760 PheThrSerGluHisTTPtTpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
Db 2345 TTTACATCAGACGACTGGCTGTGACAGATATATAAGTGAAGCACTGACACAGGGAG 2404
Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
Db 2405 ACATAGTGTCAAACTCGAGTCAGCAACATGCTCCCAACAGCAAGTATTGTCAAAG 2464
Qy 800 LysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
Db 2465 AAGACTACTAAGAGAGCGGTGCTAGCTGTTAAAATAAGTAGTGTGTTAAGAAAGCAAG 2524
Qy 820 LysIleSerLysLysThrVal 826
Db 2525 AAGACCTCTAAGAGACTGTT 2545

RESULT 5
AF132552 2417 bp mRNA linear INV 12-FEB-2003
LOCUS Drosophila melanogaster GNC1038 full insert cDNA.
DEFINITION AF132552
ACCESSION AF132552.1 GI:4689327
VERSION AF132552.1
KEYWORDS FLI CDNA.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 2417)
AUTHORS Tsang, G., Broksstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, #3200, Berkeley, CA 94720, USA
REFERENCE 2 (bases 1 to 2417)
AUTHORS Stapleton, M., Broksstein, P., Hong, L., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,

```

```

George, R., Gonzalez, M., Guarin, H., Guarin, H., Kronmiller, B., Li, P., Liao, G.,
Miranda, A., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M.
and Celniker, S.
Direct Submission
Submitted (12-FEB-2003) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.
FEATURES
source
1..2417
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
<1..2417
/genes="OstStt3"
/db_xref="FLYBASE:FBgn0011336"
<1..2395
/genes="OstStt3"
/notes="Longest ORF"
/codon_start=22
/product="Gm01938p"
/protein_id="AAD27851.2"
/db_xref="GI:28317200"
/db_xref="FLYBASE:FBgn0011336"
/translaton="LSECVAFVQIGYFRVTVGNWMDNRTPKMLNSKVAGYSSLLTF
AIIILAWAFSSRLFAVIRFESITHEFDWENYRATAYMVONGWYNLNPDERAWY
PLGRIVGTVTQGLMITSGIHLHLVNLPHIRDCVFLAPIESGLTSTSTYLTATK
ELWSACAGLPAASFAIVPGYISRSVAGSYDNEGIAIPALOPTYELWRSVKTGSVF
SAAALSYFTMSAWGGYFIINLPLPHVFLVLLIMGRISPRLLTSYFYILGLLFSM
QIBFVGQPIRTSEENAAAGVFLMLAVATLHLOSILSRNFRKLFIVGGLLVGVGV
FVAVVLTWLVGVPWPSGRFYSLMDTGYAKIHPIIASVSEHQPTTFSPFDHLILY
CAFPVGVWYCIQINDERVVVLVYASIVYFAGVNVRLMLTLPVVCMLAGVAFSGLL
DVFQEDSSKRMGTASAEATEVDEAEDSIERKTLVDKAGLKHRTKHDAQDQDYVWL
KTSIVILAVLMLLMFAVCTVWTSNAYSSIVLAFHNSODGSRNILDDPREAYWL
SQTADARVMSWDYQIOAGMANETTLVDNNTNNSHIALVGNKMSSTEEKSYEIM
TSLDYLVLVIFGVVIGYSGDDINKFLWVRIAEGEHPKDIKESDYFTDRGFRVDAE
GAPALLNLMYLSYRFGELKLDYRPGSGYDRTNRNAVIGNKDFDLTLEAYTTEHW
LVRIYRVKPHPEFRPSLTKERTIPPANFISRKNSKRRKGYIRNPVVVKGKRLK"

```

```

Alignment Scores:
Pred. No.: 2,14e-227 Length: 2417
Score: 2950.00 Matches: 557
Percent Similarity: 83.79% Conservative: 89
Best Local Similarity: 72.24% Mismatches: 113
Query Match: 72.03% Indels: 12
DB: 3 Gaps: 7
US-10-028-384-2 (1-826) x AF132552 (1-2417)
Qy 63 SerGlnProAlaGlyTTPGlnSerLeuLeuSerPheThrIleLeuPheLeuAlaThrPleu 82
Db 98 AGCAAGGTGGTGGCTACAGCAGCCTAATCACCTTCGCCATCTGCTAATACGCTGGCTG 157

```

VERSION AX799084.1 GI:37605059
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Perreault, C. and McBride, K.
 TITLE Mammalian SAMP protein, gene sequence and uses thereof in cancer therapy
 JOURNAL Patent: WO 03054008-A 3 03-JUL-2003;
 COMPATIGENE Inc. (CA)
 FEATURES
 source
 1..2710
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2710
 Score: 4263.50 Matches: 805
 Percent Similarity: 97.70% Conservative: 3
 Best Local Similarity: 97.34% Mismatch: 14
 Query Match: 96.88% Indels: 5
 DB: 6 Gaps: 2
 US-10-028-384-2 (1-826) x AX799084 (1-2710)
 QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAenSerSerProTrp 20
 Db 77 ATGGCGAGCGCTCGGGCCCGGAGAGACAGACAGAGTCTCCCTCAACTCGTCCCGTGG 136
 QY 21 SerGlyLeuMetAlaLeuGluValSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 Db 137 AGCGGCTCATGGCTCTGGGGAACAGCGCCAGCGGACCATCGGCGGAGACCCAGAGC 196
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAla---ProAlaGlyLeuSer 59
 Db 197 GCGTCCAGGGGGGGGGG-----CCGAAGCGGGGGGGGGGGGGGGGGTGTCC 244
 QY 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuSerPheThrIleLeuPheLeu 79
 Db 245 GGGGGCTGTGCGAGCGCGGGTGGAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 304
 QY 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
 Db 305 GCCTGGCTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAGAGCATATC 364
 QY 100 HisGluPheAspProTrpPheAsnTrpArgSerThrHisIleLeuAlaSerHisGlyPhe 119
 Db 365 CACGAGTTCGACCGGCTGTAACTATAGATACACATCATCTTCTCTCTCTCTCTCTCT 424
 QY 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGly 139
 Db 425 TATCAGTTTCTAAATGGTTTGAATCAAGAGAGATGGTACCCACTGGGGAAGATAGTGGT 484
 QY 140 GlyThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
 Db 485 GGCACCGTTTACCCAGGGTGTATGATTAACAGCTGGCGCTTATTCATTTGGATTTAATACA 544
 QY 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
 Db 545 TTGAACATAACAGTTCATCAAGAGATGTGTGTATTCCTTGCACCAACTTTTAAAGCGGC 604
 QY 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
 Db 605 CTTACATCCATATCATGTTCTCTGCTAAGAGACTGTGAGACCTGTGAGACCCAGAGAGACT 664
 QY 200 LeuAlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySer 219
 Db 665 CTAGCTGCTTCTCTTCT 724

220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpValLeuTrpValLys 239
 Db 725 TTTGATAATGAAGGCATTGCCATTTTGGCTTCAGTTTCTACTTACTTATGGGTAAAG 784
 QY 240 SerValLysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMet 259
 Db 785 TCTGTGAAGACCGGGTCTGTGTCTGGCAATGTGTCTGTCTGTCTGTCTGTCTGTCTGT 844
 QY 260 ValSerAlaTrpGlyGlyTrpValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 Db 845 GTCTCTGGTGGGAGGTATGTGTTCATCATCACTCACTCTCTCTCTCTCTCTCTCTCTCT 904
 QY 280 LeuLeuLeuMetGlnArgTrpSerLysArgValTrpIleAlaTrpSerThrPheTrpIle 299
 Db 905 TTGCTGTGTATGACAGAGGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 964
 QY 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 965 GTGGGTTTAAATATATTCATTCAGATACCTTTTGGGATTCAGCCANTCAGAACAGC 1024
 QY 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrp 339
 Db 1025 GAGCACATGGCAGCTGCGGTGTCTTGGCGCTGCTGCAAGCTTACGCTTTTTCAGATAT 1084
 QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 1085 CTGAGAGACCGGTTGACAAAACAGGAGTTCAGAGACCTTTCTTTTGGGTGTCTCATA 1144
 QY 360 AlaAlaGlyAlaValPheLeuSerValIleTrpLeuThrTrpGlyTrpIleAlaPro 379
 Db 1145 GCTGAGCGGCTGTGTCTCTTGTGTCTATCTATCTGACATACACAGGTATATTTGACCA 1204
 QY 380 TrpSerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIleHisIleProIle 399
 Db 1205 TGAGTGGCAGGTTTATTTTCTACTATGGGATCTGGGTATGCAAAATACACATTTCCAAT 1264
 QY 400 IleAlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHis 419
 Db 1265 ATTGCATCATGTGTCTGAACATCAGCTACGACATGGGTGTCTTTCTCTCTCTCTCTCT 1324
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 Db 1325 ATTCTTGTATGTACCTTCCAGAGCGCTATGGTCTGCAATCAAAATATCAACATGAA 1384
 QY 440 ArgValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaGlyValMetValArg 459
 Db 1385 AGAGTATTTGTCTGTATGCGATCAGTGTGTGTACTTTTGGCGAGTGTGTGTGTGTGT 1444
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 Db 1445 CTGATCTGATCTGACCCCGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1504
 QY 480 PheGluHisTrpLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSer 499
 Db 1505 TTTGAGCACTATTTGGGGATGACATGAAAGAGGAAAAACCCACCTGTGTGAGAGACAGCAGT 1564
 QY 500 AspGluAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArgLys 519
 Db 1565 GATGAGATGACAAAAGAAACCCAGGAACTTGTATGACAGGCGAGGTAAAGTGGAGAG 1624
 QY 520 HisAlaThrGlnGlnGlyThrGluGlyLeuGlyProAsnIleLysSerIleVal 539
 Db 1625 CATGTGACAGAGCAAGAGAAACCTGAGAGGGGCTTGGGCCCCCAACATCAAAAGCATTTGT 1684
 QY 540 ThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSer 559
 Db 1685 ACCATGCTGTATGCT 1744
 QY 560 AsnAlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsn 579
 Db 1745 AAGCGCTACTCAGTCCAAAGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1804
 QY 580 IleLeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAla 599

Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 774 GGCACCGTTTACCCAGGGTGGATGATAAACAGCTGGCTTATTCATTGGATTTTAAATACA 833
Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 834 TTGAACATAACAGGTTTCATTAAGAGATGTGTGTGTATTCCTTGACCAACTTTTAGCGGC 893
Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 894 CTTACATCCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAACCAACAGGAGCAGACTT 953
Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
Db 954 CTAGCTGCTTCTCATTTGCTATCGTACAGGGTACATATCTCGTCAGTGGCGGATCC 1013
Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTrpLeuTrpValLys 239
Db 1014 TTTGATAATGAAGCAATGGCAATTTTGGCTTCAGTTCACTTACTTACTTATGGGTAAAG 1073
Qy 240 SerValLysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTrpMet 259
Db 1074 TCTGTGAAGACCGGCTCTGTGTTCTGGACAAATGTCTGCTGTTCATATTTCTACATG 1133
Qy 260 ValSerAlaTrpGlyTyrValPheIleAlaLeuIleAsnLeuIleProLeuHisValPheVal 279
Db 1134 GTCTCTGGTGGGAGGTATGTGTTTCATCATCAACCTCATCCCTCTCCATGTGTGTG 1193
Qy 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTrpIle 299
Db 1194 TTGCTGCTGATGACAGAGTACAGCAGAGAGTCTACATGACATATAGCACTTTCTACATT 1253
Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db 1254 GTGGTTTAATATATATCATGCATACATACCTTTTGTGGATTTTCAGCAATCAGAACAGC 1313
Qy 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
Db 1314 GAGCACAATGGCAGCTGCGAGGTGCTTTGGCTGCTGCAAGCTTACCCTTTTTCAGATAT 1373
Qy 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
Db 1374 CTGAGAGACCGGTTGACAAAACAGCAGTTCAGACCCCTTTCTTTTGGGTGCTCACTA 1433
Qy 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
Db 1434 GCTGAGCGCTGTGTCTCTTAGTGTCTATCTGACATACACAGGTTATATGACCA 1493
Qy 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
Db 1494 TGGAGTGGCAGGTTTATCTACTATGGGATCTGGGTATGCAAAAATACACATTCCAATT 1553
Qy 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheLeuHis 419
Db 1554 ATTGCATCAGTGTCTGAACATCAGCCCTACGACATGGGTGCTTCTTCTTGTATCTACAT 1613
Qy 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
Db 1614 ATCTCTGTATGACTCTCCAGCAGGCGCTATGGTTCGTGATCAAAAATCAACAGATGA 1673
Qy 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValValValArg 459
Db 1674 AGAGTATTGTGCTCTGTATGGCATCAGTGTCTTACTTTCGGGAGTGTATGGTGGCG 1733
Qy 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
Db 1734 CTGATGCTGACTCTGACCCCGGTCGTCTGTCATGCTGTCTACTTTCGGGAGTGTATGGTGGCG 1793
Qy 480 PheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSer 499
Db 1794 TTTGAGCACTATTTGGGGGATGACATGAAGAAGGGAANAACCCACCTGTGGAGGACAGACT 1853

Qy 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
Db 1854 GATGAGGATGACAAAAGAAACCCAGGAAACTTGTGTATGACAAAGGAGGTAAAGTGGGAG 1913
Qy 520 HisAlaThrGluGlnGlyThrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
Db 1914 CATGTGACAGACGACAGGAAACCTGAGAGGGCTTGGGCCCCACATCAAAAGCATTTGTG 1973
Qy 540 ThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSer 559
Db 1974 ACCATGCTGATGCTCATGCTCTCTGATGATGTTTCGGGCTCCACTGACGCTGGGTCAAC 2033
Qy 560 AsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579
Db 2034 AAGCCTACTCTCAGTCCAGTGTGTGCTCTCTCTACATCATGATGATGATGATGATGAT 2093
Qy 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599
Db 2094 ATATTAGATGATTTTAGAAGAGCTACTTTTGGCTGGAGACAAACACCGATGAACACGCC 2153
Qy 600 ArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
Db 2154 CGGCTCATGCTGCTGGTGGACTACGGCTATCAGATTCGTCATGCCACACAGGACCACT 2213
Qy 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
Db 2214 CTGGTGGATAAACAAACACCTGGAAACACAGCCACATCGCACCTGGTCGGAAGACTATGCT 2273
Qy 640 SerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVal 659
Db 2274 TCCATGAACCGCCGCTTATAAATCATGAGGTCCCTTGAATGTCATGATGTTGTT 2333
Qy 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
Db 2334 ATTTTCGGAGAGTGTATGGCTATTCCGGGAGCATATCAACAGTTCTCTCTGGAATGTC 2393
Qy 680 ArgIleAlaGluGlyLysProLysAspIleArgGluSerAspTyrPheThrProGln 699
Db 2394 AGATAGCTGAAGGGAGCATCCCAAGACATCCGGGAAGGTGACTATTTCACCCAGCAG 2453
Qy 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLys 719
Db 2454 GGAGAGTCCGAGTAGACAAAGCTGGCTCTCTACTCTGTTAACTGCTTATGATATAA 2513
Qy 720 MetSerTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAsp 739
Db 2514 ATGTCATCTACAGATTTGGAGAAATGCTAGTATTTTCGCACTCCCGAGCTTTGAC 2573
Qy 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759
Db 2574 CGAACACGTAATGCTGAGATTGGAAATAAAGACATTAATTCACGATTTTGAGGAAGCT 2633
Qy 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
Db 2634 TTATCATCAGAGCAGCTGGCTGTGTCAGATATATAAGTGAAGACCTGACACAGGAG 2693
Qy 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
Db 2694 ACCTAGGTCAACAACCTCGAGTCACCAACATCGTCCCAACAGAGATTTTGTCAAG 2753
Qy 800 LysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
Db 2754 AAGCTACTAAAGGAAGCGTGGCTACGTTAAATATAGCTAGTGTTTAAGAAAGCGCAG 2813
Qy 820 LysIleSerLysLysThrVal 826
Db 2814 AAGACCTCTAAGGAAGACTGTT 2834

RESULT 4

AX799084

LOCUS

DEFINITION

ACCESSION

AX799084 2710 bp mRNA linear PAT 06-OCT-2003
Sequence 3 from Patent WO03054008.


```

BC052433.1 GI:30851501
MGC.
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4236)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.H., Wang, J., Hsieh, P.,
Dickenson, L., Narusnia, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, A.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, K.J., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 4236)
Strausberg, R.
Submitted (15-MAY-2003) National Institutes of Health, Mammalian
Gene Collection (MGC) Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-4590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldi, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Shier, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1. .4236
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:64679 IMAGE:6837097"
/tissue_type="Brain, enriched mouse brain 12.5dpc"
/clone_lib="NIH_BMAP_F10"
/lab_host="DH10B"
/notes="vector: pYX-ASC"
1. .4236
/gene="1300006C19Rik"
/notes="synonym: Simp"
/db_xref="LocusID:68292"

```

```

/db_xref="MGI:1915542"
366. .2837
/gene="1300006C19Rik"
/codon_start=1
/product="1300006C19Rik protein"
/protein_id="AAH52433.1"
/db_xref="GI:30851502"
/db_xref="LocusID:68292"
/db_xref="MGI:1915542"
/translacion="MAEPASPEKSHKSLNSPSWGLMALGNSRHGHGPGTQASRA
APKCPGPPAGISGGLSQACWQSLSFTILFLAMLAGSSSLPAVIFESITHEFDPM
FYRSTHDLASGFEFLNWFDERAWYELGRIVGTVPYGLMITAGLIHLINLTNIT
VHIRDVCVFLAFTFSLSTISFTLRLRELNOAGLLAACFIALIPYIGRSVAGSD
NGEIAFALQFTYILWKSVTGVSFTMCCSLFYFYMVSAGGYFIILIPHFV
LMLLMYKRVYIATSTYFIVGLISLMQIPFVQPISTSEHMAAGVALLQYAFV
QILRLTKQEPQTLFFLGSLAGAVLSVILYITGYIAPMSGRYSVLDWTGYAKI
HPIIASVSEHPTTWVSFFDLHLVCTFFAGLVFCIKINDESRVAVLVAISAVYF
AGVMYRLMLTLTPVYCMLSAFAFNVFEHVLGDMKRENPEVSDSDKRNPGNLY
DKAGVKRHVTQEPESGLGNPKSIYVLMMLMLMMPAFACTMTVTSNAYSPPSVL
ASYNHDTNILDFFREAYFWLRQNTDSEHARVMSWVDYQIAGMAHRTILVDNNTWN
NSHIALVGMSSNETAAYKINRSLDYLIVFGVIGYGGDDINKFLMWVIAEGE
HPKIDREGDYFTQQGEFRVDKAGSPTLLNCLMYKSYRFGEMOLDFTPPGDRTN
AIGNKDIKFKHLEAFTSEHVLVRIYKKAAPDNRETILGHKPRVTNIVPKQKLSKKT
TKRKGIVKVKLVFKKTKTKTV"
567. .2537
/gene="1300006C19Rik"
/notes="STT3; Region: Oligosaccharyl transferase STT3
subunit. This family consists of the oligosaccharyl
transferase STT3 subunit and related proteins. The STT3
subunit is part of the oligosaccharyl transferase (Oase)
complex of proteins and is required for its activity.
Oase transfers a lipid-linked core-oligosaccharide to
selected asparagine residues in the ER"
/db_xref="CDD:pfam02516"

```

misc_feature

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

Length:

Matches:

Conservative:

Mismatches:

Indels:

Gaps:

US-10-028-384-2 (1-826) x BC052433 (1-4236)

Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTTP 20

Db 366 ATGGCGGAGCCCTCGGCCCGGAGAGCAAGCACAGTGTCTCCCTCACTCGTCCCGCTGG 425

Qy 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40

Db 426 AGCGGCTCATGCTCTCGGGAAACAGCCGCCACGACCATGGCCCGGACCCAGCAGC 485

Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAla---ProAlaGlyLeuSer 59

Db 486 GCGTCCAGCGCGCGCGCGG-----CGAAGCGCGCGCGCGCGCGCGCGCTCC 533

Qy 60 GlyGlyLeuSerGlnProAlaGlyTTPGlnSerLeuLeuSerPheThrLeuLeuPheLeu 79

Db 534 GGGGGCTTGTTCGACGCGCGCGGGTGGCAGTGTGTCTCTCTTCCATCATCTCTCTCT 593

Qy 80 AlaTTPLeuAlaGlyPheSerSerArgLeuPheAlaValIleAtpGheGluSerIleIle 99

Db 594 GCCTGGCTGGCGGCTTCAGCTCGCGCTCTTCGCGGTATCTCCGCTTCGAGAGCATCATC 653

Qy 100 HisGluPheAspProTTPPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119

Db 654 CACGAGTTCGACCGCGGTAACTATAGATCAACACATCATCTTGCATCTCTCATGAGTTC 713

Qy 120 TyrGluPheLeuAsnTTPPheAspGluArgAlaTTPVrProLeuGlyValGlyValGly 139

Db 714 TATGAGTTCTTAATTTGTTTGTGTAAGAGCATGGTACCCACCTGGGAAGATAGTGGGT 773

QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
DB 421 ACTGTTTACCCAGGGTTGATGATTAACCGCTGGCTTATTCATTGGATTTTAATACATTG 480
QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
DB 481 AACATAACTGTTTCATAGAAGAGAGCTATGCTGTTCTTCACCAACTTTTAGCGGCTT 540
QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
DB 541 ACATCTATATCTTCTCTCTTACAGAGAACTTTGGAACCAAGAGAGACGACTTTTA 600
QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
DB 601 GCTGCTGTTTATTGCTTATGTACAGGCTACATATCTCGGTCACTAGCTGATCTCTT 660
QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrIleTrpValIleSer 240
DB 661 GATAATGAAGGCAATGCTTATTTTGGCACTTCAGTTCACATCTATTTATGGGTAATCT 720
QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
DB 721 GTAAAACTGGGTCACTTTTGGCAATGTGCTGCTTATCTTATCTATATGCTG 780
QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
DB 781 TCTGCTTGGGTGGTATGATTTATCATCAATCTTATCCACTGATGATTTTGGTGG 840
QY 281 LeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
DB 841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTTCACTTGTG 900
QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
DB 901 GGTTTAAATATATCAATGCAATACCTTTTGTGGGATTTCCAGCAATACAGCAAGTGAA 960
QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
DB 961 CACATGGCAGCTGCAGGTGCTTTTGGCATTTGCTGCAAGCTTATGCTTTCTTGCAGTATCTG 1020
QY 341 ArgAspArgLeuThrLysGlnIlePheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
DB 1021 AGAGACCATTAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCAGTACCT 1080
QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrp 380
DB 1081 GCAGGTGCTGTCTTCTTGTGTCATCTATTGACTTATACAGGTTACATTCACCATCG 1140
QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIle 400
DB 1141 AGTGCAGGTTTATTCATTGTTGGATCTGGGTATGCAAAATATACATTCCTCAATATT 1200
QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
DB 1201 GCATCAGTGTCTGAGCATCAACCTACGACTGGGTGCTTCTTCTTGTATCTACATATT 1260
QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
DB 1261 CTGTGATGACTTCCCGAGCGGCTTTGGTTCTGCATCAAAATATCAACGATGAAGA 1320
QY 441 ValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeu 460
DB 1321 GPATTTGTTGCTATATGCAATCAGTGTCTCTTCTTCTGAGAGTATGGTGGCAGCTG 1380
QY 461 MetLeuThrLeuTrpProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
DB 1381 ATGTTGACTTTGACTCCAGCTGTTGATGCTGTCTCAATTCCTTTTCAAAATGTTTT 1440
QY 481 GluHisTyrLeuGlyAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
DB 1441 GAGCACTATTTGGGGATGATGATGAAAGGGAATCCACTGCTGGAGGACAGCAGTAT 1500
QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHis 520

DB 1501 GAGGATGACAAAAGAACCAAGGAATTTGTATGATAAGCGAGGTAAAGTGAGGAAACAT 1560
QY 521 AlaThrGluGlnGlyThrGluGlyLeuGlyProAsnIleLysSerIleValThr 540
DB 1561 GCRAACTGAACAGGAAACAACTGAAGAGGATTAGGCCCTTAATATAAAGCATTTGTACC 1620
QY 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
DB 1621 ATGTTGATGCTGATGCTATTGATGATTTGCTGCTCCACTGTACCTGGGTCAACAGCAAT 1680
QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
DB 1681 GCCTACTAGTCCAAAGTGTAGTCTCTGGCTCATCAATCATGATGCGCACCAAGGAATATC 1740
QY 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
DB 1741 TTATGATGATTTTAGAAGAGCTTACTTTTGGCTTAAGCAAAATATACAGATGAACATGCAGA 1800
QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
DB 1801 GTAATGCTTGTGGGATTTATGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860
QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaValMetSerSer 640
DB 1861 GTGATATATTAACACCTGGGAATACACCCACATAGCACCTGTGGGAAAGACTATGTTCT 1920
QY 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
DB 1921 AATGAAACAGCAGCTTATAAATCATAGGACTCTAGATGTAGATTAATGTTTGGTTATT 1980
QY 661 PheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
DB 1981 TTTGGAGGGTTTATGGCTATTTCTGGTGATGATATCACAAAATTTCTCTGGATGTTTAG 2040
QY 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
DB 2041 ATAGTCAAGGAGAACATCCAAAGACATTCGGAAAGTACTATTTTACCCACAGGGA 2100
QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
DB 2101 GAATTCGGTTAGCAAAAGCAGGATCCCTACTTTTGTGTAATGCTTATGATATAAATG 2160
QY 721 SerTyrTrpArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
DB 2161 TCATCTACAGATTTGGAGAAATGCAGCTGGATTTTTCGTACACCCCGAGGTTTTCACCGA 2220
QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
DB 2221 ACAGTAACTGAGATTTGAAATAAGGACATTAATTCAAACATTTTGAAGAAGACCTTT 2280
QY 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
DB 2281 ACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGACACCTGATAACAGGAGACA 2340
QY 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
DB 2341 TTATGATCAAAACCTCGATCCACCACTTTTCCCAAAACAGAGATATTGTCACAAAGAG 2400
QY 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysGlyLysLys 820
DB 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAAGCTGGTTTAAAGAAAGGCAAGAAA 2460
QY 821 IleSerLysLysThrVal 826
DB 2461 ATAATCTAAGAAGACTGTT 2478

RESULT 3

BC052433

LOCUS

DEFINITION

MUS MUSCULUS RIKEN CDNA 1300006C19 GENE, mRNA (CDNA CLONE MGC:64679 IMAGE:6837097), COMPLETE CDS.

ACCESSION

BC052433

4236 bp

mRNA

linear

ROD 12-NOV-2003

QY 601 ValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaSerArgThrIleu 620
 Db 1801 GTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGAATGCTAATAGAACTAGCTTTG 1860
 QY 621 ValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 Db 1861 GTGGATATACACCTGGATACAGCAATAGCACTGGTGGAAAGCTATGCTCTCT 1920
 QY 641 AsnGluThrAlaIleLysIleMetArgThrLeuAspValAspTyrValleuValle 660
 Db 1921 AATGAACAGCAGCAGCTATAAATCATGAGACTCTAGATGATGATTTTGGTTATT 1980
 QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
 Db 1981 TTTGGAGGGTATTGGCTATTCTGGTGTATGATATCAACAAATTTCTGGATGGTAGG 2040
 QY 691 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGA 2100
 QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 2101 GAATTCCTGTAGCAAGCAGGATCCCTACTTTTGTGATTCGCTTATGATAAATG 2160
 QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
 Db 2161 TCATACACAGATTGGAGAAATGCAGCTGGATTTTTCGTACACCCCGAGTTTTCACCGA 2220
 QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 Db 2221 ACACGTATGCTGAGATGGGATTAAGCAATTAATTCAACTTTTGAAGAGCCCTT 2280
 QY 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2281 ACATCAGAACACTGGCTTTGTAGGATATATAAAGTAAAGCACCTGATACAGGAGACA 2340
 QY 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2341 TTAGATCAACAACCTCGAGTCACCAACATTTCCCAACAGAGATTTTGTCAAGAGAG 2400
 QY 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 Db 2401 ACTACCAAGAGAGCGCTGCTACATTAATAAATAAGCTGTTTTTAAAGAGCAAGAAA 2460
 QY 821 IleSerLysLysThrVal 826
 Db 2461 ATAATCAAGAGACTGTT 2478

RESULT 2
 AY074880
 LOCUS
 DEFINITION Homo sapiens source of immunodominant MHC-associated peptides
 (STMP) mRNA, complete cds.
 ACCESSION AY074880
 VERSION AY074880.1 GI:19879588
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 2481)
 McBride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A.,
 Bergeron, J., and Perreault, C.
 TITLE The model B6dom1 minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast SRT3 gene
 JOURNAL Immunogenetics 54 (8), 562-569 (2002)
 MEDLINE 2326278
 PUBMED 12439619
 REFERENCE 2 (bases 1 to 2481)
 McBride, K. and Perreault, C.
 TITLE Direct Submission
 AUTHORS
 Submitted (22-JAN-2002) Molecular Biology Group, Compattigene, 6100
 Royalmount, Montreal, Qc H4P 2R2, Canada

FEATURES

source

Location/Qualifiers
 1..2481
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 1..2481
 /gene="STMP"
 1..2481
 /gene="STMP"
 /note="similar to Saccharomyces cerevisiae OSTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018759; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"

/protein_id="AAL71884.1"
 /db_xref="GI:19879588"
 /translation="MAEPSPAPESKHSLSNPSWGLMAGNSHHGHGPAQCAHKA
 AGCAAPPKAPAGLSQSPAGMCSLSTILFLAMAGFSSSLFAVIFESIIHF
 DWFNVRSTHLASHGFYELFNWDERAWYPLGRIVGTVPGLMIFAGLHMLNTL
 NTVHIDVCVFLAPTSLSISTELTRELNOAGLLAACFIAIVPGVILSRVAG
 SDNGLAIFALQFTYVLWKSVKTSVFTWCCLSVFYMVSAWGGVFLINLPLH
 VFLLMGRYSKRYIAYSTFYVGLILSLMQIPFVGPQPIRTSEHMAAAGVALLQAY
 AFQIDRLTKQERQTLFLGSLAGAVLSVILTYTGTITAPWSGRYSVLDQY
 AKIHPIIASVSEHQFTTWVFFDFDLHLCTFFPAGLWFCIKINIDERSVALYISA
 VYFAGVMVRLMLTLTPVCMLSAIFSNVPHYGLDDMKRENPEVDSDEDDKRNQ
 NLYDAGKVRKHAITEQETEEGLGPKSIIVTMLMLLMMFAVHCTVTSNAYSPP
 TVNASHIALGKAMSSNETAAKIMRLTDVYLVIFGVYIGVDDINKFLMMVRIA
 EGEHPKIDRESYFTPQEFVDKAGSPILLNCLMYKMYRFGEMOLDERTPPQDR
 TRNAGIKNDIKFKLEAEFTSEHLVRIYKVKAPDNRETLIDHKPRVTNIFPKQKLS
 KKTTRKRKYINKLVFKKKKSKKTV"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2481
 Score: 4401.00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x AY074880 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
 Db 1 ATGGCGAGCCCTCGGCCCGGAGAGCAAGCAAGTCTCTCTCACTCGTCCCGCTGG 60
 QY 21 SerGlyLeuMetAlaGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
 Db 61 AGTGGCCTCATGCGCCCTGGGAAACAGCGGCACGCCACCGCGCGCGCGCCAGTGC 120
 QY 41 AlaHisLysAlaGlyGlyValAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 Db 121 GCGCACAGCGG 180
 QY 61 GlyLeuSerGluProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 Db 181 GGGCTGTGCGAGCGCGCTGGTGGCAGTGGCTCTCTCTCTTCCACCATCTCTTCTCGGCC 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleAspPheGluSerIleIleHis 100
 Db 241 TGGCTTGGCGCTTCACTCGCGCTCTTGGCCGTCATCCGCTTCGAAAGAGATCATCCAC 300
 QY 101 GluPheAspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
 Db 301 GAGTTCGACCCCGTGGTGTAACTATAGATCAACACATCATCTTGTGATCTCATGGTCTCTAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 140
 Db 361 GAATTTTAAATTGGTTTGTATGATAAGAGCATGGTATCCACTAGGAGATAGTAGTGGT 420

AGGAAPKPPAPAGLSGLSOPAGWQSLLSFTILFLAWLAGFSRLFAVIRPESIIHF
DPWFYRSTHLSHSGYEFELNWFDERAWTPLGRIVGGTVIPGLMTAGLIHWLNTL
NITVIRDCVFLAPTSGLTSISFTLLTHLWQAGLLAACFIAIPVGYISRSVAG
SFDNEGAIYALQYLYLWVKSFKTSFVWCCLSYFTWAMGGYFIINLIPLH
VFLLLMQRYSKQVYIAYSTFYIVGLILSMQIPFVGFQPIRTSEHMAAGVALLQAY
AFLOYDRDRITKQBFYFLGLSLAAGAVLSVILVTYGYIAPSGRFRVSLMDTGY
AKIHIPIIASVSEHQPWTWVSFFDLHLVCTPAGLWFCIKNIDRFRVVALISA
VYPAGVMVRLMLTLPVCMLSALAFSNVFEHLGDDMKRENPPVEDSDEDDKRNQ
NLVDKACKVRKHATEQKTEGELGPNIKSIIVTMLMLLWMPFAVCHTWTISNAYSFS
VFLASYNHEDGTRNLLDFREAYFWLRQNTDEHARVMSMDYGOIAGMARNTLLVDNN
TWNNSHLVKGKMSNETAAYKIMRTLDYVIVLIVFGVYGGSDINFKFLMWYIA
EGEHPKDIRESDYFTPOGEFRVDPKAGSPTLLNCLMYKMSYREFGEMQLDFRPPGDR
TRNABIGNKDIKFKHLEBEFTSEHLVRIYVKRAPDNRETLDHKPRVTNIPFKQKLYS
KKTITKRGVIKKNLVFKGKKISKXTV"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2481
Score: 4401.00 Matches: 826
Percent Similarity: 100.00% Conservations: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x AX799082 (1-2481)

Qy 1 MetAlaGluProSerAlaProGluSerLysHisSerLeuAsnSerSerProTrrp 20
Db 1 ATGGCGGAGCCCTCGGCGCGGAGGAGCAAGTCGTCCTCAACTCGTCCCGTGG 60
Qy 21 SerGlyLeuMetAlaLeuGlyAanSerArgHisGlyHisGlyProGlyAlaGlnCys 40
Db 61 AGTGGCCCTCATGGCCCTTGGGAAACAGCGGCGACCGCCACCGCGCGCGCCAGTGC 120
Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
Db 121 GCGCACAGCGGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Qy 61 GlyLeuSerGlnProAlaGlyTrrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
Db 181 GGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Qy 81 TrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
Db 241 TGCGCTTGGCGGCTTCAGTCGCGCTTCTCGCGCTCATCGCTTCGAAAGCATATCCAC 300
Qy 101 GluPheAspProTrrpPheAsnTrrpArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
Db 301 GAGTTCGACCGGTGTTAACTATAGATCAACACATCATCTTGCATCTCATGGTTCAT 360
Qy 121 GluPheLeuAsnTrrpPheAspGluArgAlaTrrpTrrpProLeuGlyArgIleValGlyGly 140
Db 361 GAATTTTAAATTTGGTTTATGAAGAGCATGTATCCACTAGGAAGATATGATAGTGGT 420
Qy 141 ThrValTrrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrrpIleLeuAsnThrLeu 160
Db 421 ACTGTTTACCCAGGTTGATGATAACCGTGGCTTATTCATGGATTTTAAATACATG 480
Qy 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
Db 481 ACATTAACGTGTTACATAAGACAGCATGTGTGTCTTGCCTTGCACCAACTTTTAGCGGCT 540
Qy 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrrpAsnGlnGlyValaGlyLeuLeu 200
Db 541 ACATCTATATCTACTTCTGCTTACAGAGAACTTTGGAAACAGGAGGAGGAGCTTTTA 600
Qy 201 AlaAlaCysPheIleAlaIleValProGlyTrrpIleSerArgSerValAlaGlySerPhe 220
Db 601 GCTGCTGTGTTTATGCTATTGATCCAGGCTACATATCTCGGTAGTGTAGTCCCTTT 660
Qy 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrrpTrrpValIlySer 240
Db 661 GATATGAGGCGATGCTATTTTGGCACTTCAGTTCACATACTATTTATGGGTAAATCT 720

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 9150.76 Seconds
(without alignments)
4268.840 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 4401
Sequence: 1 MAEPAPESKHKSLSNPSW.....GVKNKLVFKGKISKTKTV 826

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app.query.fasta_1.3740
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOCL=0 -LOOEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US10028384 @CGN 1.1 20262 @runat_14122004_131519_8017 -NCFU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 4401 | 100.0 | 2481 | 6 | AX799082 Sequence |
| 2 | 4401 | 100.0 | 2481 | 9 | AX074880 Homo sapi |
| 3 | 4269.5 | 97.0 | 4236 | 10 | BC052433 Mus muscu |
| 4 | 4263.5 | 96.9 | 2710 | 6 | AX799084 Sequence |

| | 5 | 2950 | 67.0 | 2417 | 3 | AF132552 |
|---|----|--------|------|--------|----|-------------------|
| | 6 | 2950 | 67.0 | 2417 | 6 | AX799088 |
| | 7 | 2950 | 67.0 | 2699 | 6 | CQ589353 Sequence |
| C | 8 | 2838 | 64.5 | 4922 | 6 | CQ589352 Sequence |
| | 9 | 2838 | 64.5 | 162321 | 3 | AC007853 |
| C | 10 | 2838 | 64.5 | 181132 | 3 | AC008206 |
| C | 11 | 2838 | 64.5 | 227219 | 3 | AE003750 |
| C | 12 | 2791.5 | 63.4 | 75650 | 2 | AC018145 |
| | 13 | 2678 | 60.8 | 1734 | 10 | BC003206 |
| | 14 | 2605 | 59.2 | 54118 | 3 | U13019 |
| | 15 | 2370 | 53.9 | 1664 | 6 | CQ782554 |
| | 16 | 2370 | 53.9 | 1664 | 6 | BD127193 |
| | 17 | 2370 | 53.9 | 1664 | 9 | AK074587 |
| | 18 | 2274 | 51.7 | 2855 | 6 | CQ573675 |
| | 19 | 2266.5 | 51.5 | 3094 | 6 | AX799090 |
| | 20 | 2266.5 | 51.5 | 3094 | 10 | MUSTRPR |
| | 21 | 2266 | 51.5 | 2736 | 10 | BC037612 |
| | 22 | 2262 | 51.4 | 2118 | 9 | BT007100 |
| | 23 | 2262 | 51.4 | 2118 | 12 | BT008132 |
| | 24 | 2262 | 51.4 | 2450 | 9 | BC048348 |
| | 25 | 2262 | 51.4 | 2458 | 6 | CQ834714 |
| | 26 | 2262 | 51.4 | 2472 | 6 | CQ727867 |
| | 27 | 2262 | 51.4 | 2516 | 9 | BC020965 |
| | 28 | 2260 | 51.4 | 2293 | 9 | HUMTMC |
| | 29 | 2259 | 51.3 | 2608 | 5 | BC046072 |
| | 30 | 2258.5 | 51.3 | 3686 | 10 | AK129027 |
| | 31 | 2256 | 51.3 | 2481 | 5 | BC067313 |
| | 32 | 2255 | 51.2 | 2898 | 9 | HS0809254 |
| | 33 | 2253 | 51.2 | 2567 | 5 | BC063234 |
| | 34 | 2250 | 51.1 | 2472 | 6 | AX799092 |
| | 35 | 2250 | 51.1 | 2472 | 9 | HUMBSA |
| | 36 | 2240.5 | 50.9 | 2435 | 6 | CQ834710 |
| | 37 | 2236 | 50.8 | 2760 | 6 | CQ42624 |
| C | 38 | 2188.5 | 49.7 | 110000 | 8 | CR382128_21 |
| | 39 | 2171.5 | 49.3 | 2555 | 6 | BT012881 |
| | 40 | 2164 | 49.2 | 2256 | 6 | AX489023 |
| | 41 | 2147.5 | 48.8 | 110000 | 8 | CR382134_03 |
| | 42 | 2138.5 | 48.4 | 2756 | 8 | AK099138 |
| | 43 | 2123 | 48.2 | 2819 | 8 | AK067609 |
| | 44 | 2103.5 | 47.8 | 300029 | 8 | AE016903 |
| C | 45 | 2096 | 47.6 | 110000 | 8 | CR382121_09 |

ALIGNMENTS

RESULT 1
AX799082
LOCUS AX799082 2481 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 1 from Patent WO03054008.
ACCESSION AX799082
VERSION AX799082.1 GI:37605057
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. Perreault, C. and McBride, K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
Patent: WO 03054008-A 1 03-JUL-2003;
COMPATIGENE INC. (CA)
FEATURES
Location/Qualifiers
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2481
/codon_start=1
/product="SIMP"
/protein_id="CAE48447.1"
/db_xref="GI:37605058"
/translation="MAEPAPESKHKSLSNPSWGLMALGNRHHGHGFGAQAHA"

cdna Library Preparation: M. Bento Soares, University of Iowa
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDAM602 row: i column: 23
 High quality sequence stop: 594.

FEATURES

source

Location/Qualifiers

1..709

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:30523894"

/tissue_type="Pooled Chondrosarcoma Tumor cells"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_219"

/note="Vector: pX-Asc; Site 1: EcoRI; Site 2: NotI;

Library is oligo-dT primed and directionally cloned.

Denatured RNA was size fractionated on a 1% agarose gel.

First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to RNA size fraction, ligated with

EcoR I adaptor, digested with Not I and then cloned

directionally into pX-Asc vector. Average insert size

0.5-1kb. Adaptors 5' (AATTCGGCAGG)3' and 5'd

(CTCTGGCG)3'..3' linker sequence - GCGGCGCTGAGGCC T18.

Sequencing primers 3' end: T3 promoter primer 5'd

(ATTAAACCTCACTAAAGGA)3'. 5' End: T7 promoter primer 5'd

(TATACCACTCACTAAGG)3'. Library was constructed in the

laboratory of M. Bento Soares. Average insert size 2-3kb.

Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 4,43e-208 Length: 709
 Score: 203.00 Matches: 203
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatch: 0
 Query Match: 24.58% Indels: 0
 DB: 7 Gaps: 0

US-10-028-384-2 (1-826) x CF780700 (1-709)

Qy 608 GlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsn 627
 Db 9 GGCATCAGATCGCTGGAATGGCTAATAGAACTACGTTGGTGATTAATACACCTGGAAT 68
 Qy 628 AsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLys 647
 Db 59 AACAGCCACATAGACCTGGTGGGAAAGCTATGCTCTTCTAATGAACACGACCTATAAA 128
 Qy 648 IleMetArgThrLeuAspValAspTyrValIleValIlePheGlyGlyValIleGlyTyr 667
 Db 129 ATCATGAGGACTCAGATGTAGATTATGTTTGGTTATTTTGGAGGGGTATTGGCTAT 188
 Qy 668 SerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlyGluHisPro 687
 Db 189 TCTGGTATGATATCAACAATTTCTCTGGATGGTTAGATAGCTGAGGAGAAATCC 248
 Qy 688 LysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAla 707
 Db 249 AAAGACATTCGGGAAGTGAATTTTACCCACAGGAGAAATTCGCTAGACAAAGCA 308
 Qy 708 GlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGlu 727
 Db 309 GGATCCCTACTTGTGTAATTTGGCTTATGTAATAAATGTCACTACAGATTTCGAGAA 368
 Qy 728 MetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGly 747
 Db 369 ATGAGCTGGATTTCGTACACCCCGAGGTTTTCACCGAACACGTAATGCTGAGATTGGA 428
 Qy 748 AsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuVal 767

Db 429 AATAAGGACATTAAATTTCAACATTGGGAAGGCTTTATCATCAGAACACTGGCTTGT 488
 Qy 768 ArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgVal 787
 Db 489 AGGATATATAAAGTAAAGACCTGATACAGGGAGACATTAGATCACAACCTCGAGTC 548
 Qy 788 ThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGly 807
 Db 549 ACCAATTTTCCAAAACAGAAATTTTGTCAAGAAGACTTACCAAAAAGGAGCGTGGC 608
 Qy 808 TTTTLeLys 810
 Db 609 TACATTAAA 617

Search completed: December 15, 2004, 06:13:58
 Job time : 6420.05 secs

RESULT 14
CF893729
LOCUS
DEFINITION
620 bp mRNA linear EST 04-NOV-2003
A0126H08-5 NIA Mouse Undifferentiated ES Cell cDNA Library (Long 1)
Mus musculus cDNA clone NIA:A0126H08 IMAGE:30726139 5', mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
Piao Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method

JOURNAL
MEDLINE
PUBMED
COMMENT
Genome Res. 11 (9), 1553-1558 (2001)
21429098
11544199
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: A0126 row: H column: 08
Seq primer: M13 Reverse
High quality sequence stop: 620
POLYA=No.

FEATURES
source
Location/Qualifiers
1..620
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129/Sv x 129/Sv-CP"
/db_xref="niaEST:A0126H08-5"
/db_xref="taxon:10090"
/clones="NIA:A0126H08 IMAGE:30726139"
/dev_stage="R1 ES cells"
/lab_host="DH10B"
/clone_lib="NIA Mouse Undifferentiated ES Cell cDNA
Library (Long 1)"
/note="vector: pCMV-SPORT6 (Invitrogen); site 1: SalI;
site 2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001)). [PMID: 11544199]. Total
RNAs were obtained from Dr. Kenneth R. Boheler (National
Institute on Aging, USA). ES cells were cultured without
feeder cells in the presence of LIF and BRL-conditioned
media. Double-stranded cDNAs were synthesized with an
Oligo(dT) primer (Invitrogen):
5'-TGACTAGTCTAGATCGAGCGCGCCCTTTT-3' from
14.2 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lene-linker L1-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.4 kb. The library was
constructed by Yulan Piao."

ORIGIN

Alignment Scores:
Pred. No.: 2,2e-211 Length: 620
Score: 206.00 Matches: 206

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 24.94% Indels: 0
DB: 7 Gaps: 0

US-10-028-384-2 (1-826) x CF893729 (1-620)

QY 260 ValSerAlaTrpGlyGlyTyrValPheIleLeuLeuPheVal 279
DB 2 GTCTCTCGTGGGAGGTTATGTTTCATCATCATCCCTCTCATGTGTGG 61
QY 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
DB 62 TTGCTGCTGATGTCAGAGGTACAGCAAGAGAGTCTACATAGCATATAGCACTTCTACATT 121
QY 300 ValGlyLeuLeuLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
DB 122 GTGGGTTTAAATATATTCATTCAGAGTACCTTTTGGGATTTTCAGCAATCAGAACAGC 181
QY 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
DB 182 GAGCACATGTCAGCTGCAGGTGCTTTTGGCTGCTGCAAGCTTAGCTTTTTCAGTAT 241
QY 340 LeuArgSerPheLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
DB 242 CTGAGAGACCGGTGACAAACAGAGGTCCAGACCTTTTCTTTTGGGTCTTCACAT 301
QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
DB 302 GCTCAGCGCTGTGTTCTTCTAGTGTCTATCTGACATACACAGGTATATATTGCACCA 361
QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaValIleHisIleProIle 399
DB 362 TGGAGTGGCAGGTTTATTCTATGCGATCTGGGTATGCAAAATACACATTCCAATT 421
QY 400 IleAlaSerValSerGluHisGlnProThrTrpValSerPhePhePheLeuHis 419
DB 422 ATTGCATCAGTGTCTGAACATCAGCTCAGCATGAGGTGCTTCTTCTTTCATCTACAT 481
QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleIleAsnIleAsnAspGlu 439
DB 482 ATTCTTGATGTACCTTCCAGCGCCCTATGTTCTGCATCAAAATATCAACATGNA 541
QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
DB 542 AGAGTGTTCGCTGTATGCGATCAGTGTCTTCTACTTTGCCGAGTGTATGTCGG 601
QY 460 LeuMetLeuThrLeuThr 465
DB 602 CTGATGCTGACTCTGACC 619

RESULT 15

CF893729
LOCUS
DEFINITION
709 bp mRNA linear EST 20-OCT-2003
AGENCOURT 15873364 NIH MGC 219 Homo sapiens cDNA clone
IMAGE:30523894 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

CF893729
CF893729.1 GI:37740297

ORGANISM
Homo sapiens (human)

REFERENCE
1 (bases 1 to 709)

1 (bases 1 to 709)
NH-MGC <http://mgs.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: csapsb-xemail.nih.gov
Tissue Procurement: James Martin, University of Iowa

feeder cells in the presence of LIF and BRL-conditioned media. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]:
 5'-pGACTAGTCTTAGATCGGCGCGCCCTTTTTTTTTT-3' from 14.2 ug of total RNA, treated with Ta DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lons-linker Lu-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yulan Piao."

ORIGIN

Alignment Scores:
 Pred. No.: 5,11e-220 Length: 645
 Score: 214.00 Matches: 214
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.91% Indels: 0
 DB: 7 Gaps: 0

US-10-028-384-2. (1-826) x CF893672 (1-645)

QY 260 ValSerAlaTrpGlyGlyTyrValPheIleAsnLeuIleProLeuHisValPheVal 279
 Db 2 GTCCTGCGTGGGAGGTATGTTTCATCATCACTCATCCCTCCCTCATGTTG 61
 QY 280 LeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIle 299
 Db 62 TTGCTGCTGATGCGAGGTACAGCAGAGTCTTACATAGCATATAGCACTTCTACAT 121
 QY 300 ValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db 122 GTGGGTTTAAATATTATCATCATCACTTCTTGTGGATTTTCAGCAATCATCAAGC 181
 QY 320 GluHisMetAlaAlaGlyValPheAlaLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 Db 182 GAGCAGATGGGAGCTGCGAGTCTTTCGCTGCTGCAAGCTTACCTTTTTCAGTAT 241
 QY 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeu 359
 Db 242 CTGAGAGACCGGTTGACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGCTCACAT 301
 QY 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
 Db 302 GCTGAGCGCTGTGTTCTTCTAGTGTCTATCTGACATACACAGGTATATATGACCA 361
 QY 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
 Db 362 TGGAGTGCAGGTTTATCTACTATGGATCTGGTATCGGATCGAATAATACATCCATT 421
 QY 400 IleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePhePheLeuHis 419
 Db 422 ATTGCATCAGTGTCTGAACATCACCCCTACGACATGGGTGCTTCTTCTTGTATCAT 481
 QY 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 Db 482 ATCTCTGTATGATCTTCCAGCAGGCTATGTTCTGTCATCAAAAATATCAACGATGA 541
 QY 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 Db 542 AGAGTGTTTGTGCTGCTGTATGCGCATCAGTGTCTTCTTCTTCTTCTTCTTCTTCT 601
 QY 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAla 473
 Db 602 CTGATGCTGACTCTGACCCCGGTGCTGTCATGCTGTGCGCC 643

RESULT 12

CF742714
 LOCUS
 DEFINITION
 UI-M-HBO-clm-m-17-0-UI.r1 NIH BMAP_HBO Mus musculus cDNA clone
 IMAGE:30618064 5', mRNA sequence.
 CF742714
 ACCESSION
 VERSION
 CF742714.1 GI:37639053
 EST.
 KEYWORDS
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 692)
 REFERENCE
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/mousefl.html
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

FEATURES
 Location/Qualifiers
 1..692
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30618064"
 /tissue_type="whole eye"
 /dev_stage="embryo 12.5,13.5,14.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP_HBO"
 /notes="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TTTATGAGCT. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
 Developing Mouse Nervous System", supported by National
 Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:
 Pred. No.: 5,5e-220 Length: 692
 Score: 214.00 Matches: 227
 Percent Similarity: 99.56% Conservatives: 0
 Best Local Similarity: 99.56% Mismatches: 0
 Query Match: 25.91% Indels: 1
 DB: 7 Gaps: 0

US-10-028-384-2 (1-826) x CF742714 (1-692)

QY 198 GlyLeuLeuAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValala 217
 Db 2 GGATCTTCTAGCTGCTTCTTCTATGCTGCTACCGGATCATATCTCGGTCAGTGGG 61
 QY 218 GlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrp 237
 Db 62 GGATCCCTTGTATTAATGAAGGCATTCGCAATTTTGGCTTCAGTTCATCTACTATTAG 121
 QY 238 ValLysSerValLysThrGlySerValPheThrMetCysCysLeuSerTyrPhe 257

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 National Cancer Genomics
 Office of Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Stefan Hanson
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
 and advice from Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: NDA615 row: k column: 08
 High quality sequence stop: 625.

FEATURES source
 1. .861
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30528919"
 /tissue_type="Human Placenta"
 /lab_hosts="DH10B TorA"
 /clone_lib="NIH_MGC 147"
 /note="Organ: placenta; Vector: pBluescriptR; Site: 1:
 all-XhoI; Site 2: BamH; Oligo-dT primed using primer
 5'-TTTTTTTTTTTTTNN-3', size-selected for average
 insert size 2.3 kb and normalized to 5. This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein
 (NIH/NHGRI, National Institutes of Health). Note: This is
 a NIH_MGC library."

ORIGIN
 Alignment Scores:
 Pred. No.: 2,69e-225 Length: 861
 Score: 219.00 Matches: 232
 Percent Similarity: 99.57% Conservative: 0
 Best Local Similarity: 99.57% Mismatches: 0
 Query Match: 26.51% Indels: 1
 DB: 7 Gaps: 0

US-10-028-384-2 (1-826) x CF995035 (1-861)
 QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
 DB 27 GCCTACTCTAGTCCAGGTAGTCTGCTGCTCATAAATCATGATGCCACAGGAATATC 86
 QY 581 LeuAspAspPheArgGluAlaTyrPheTyrLeuArgGlnAsnThrAspGluHisAlaArg 600
 DB 87 TTAGATGATTTTACAGAGCTTACTTTTGCTAAGCAAAATACAGATCAATGCACGA 146
 QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 DB 147 GTAATGCTTGTGGGATTATGGCTATCATAGATAGCTGGAATGGCTAATAGAACTACGTG 206
 QY 621 ValAspAsnAnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 DB 207 GTGGATAATAACACCTCGAATAACAGCCATAGCACTGGTGGGAAACATATGCTTCT 266
 QY 641 AsnGlnThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValleuValle 660
 DB 267 AATGAAACAGCAGCCTATAAAATCATGAGACCTAGATGTAGATTATTTTGTGTTATT 326
 QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
 DB 327 TTTGGAGGGGTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCGATGGTATGG 386
 QY 681 IleAlaGluGlyGluHisPheLysAspIleArgGluSerAspTyrPheThrProGly 700

Db 387 ATAGCTGAAGGAGACATCCCAAGACATTCGGAAAGTACTATTTTACCCACAGGA 446
 QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrIleMet 720
 Db 447 GAATTCGGTGTAGACAAAGCAGGATCCCTACTTTGTTGAAATTCCTTTATGATAAAATG 506
 QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
 Db 507 TCATACACAGATTGGAGAAATGCAGCTGGATTTCGTACACCCAGGTTTGTACCGA 566
 QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 Db 567 ACAGCTAATGCTGAGATTGGAATAAGGACATTAATAATCAACATTTGGAAGAGCCTTT 626
 QY 761 ThrSerGluHisTrpLeu-ValArgIleTyrLysValLysAlaPheAspAsnArgGluTh 780
 Db 627 ACATCAGAACACTGGCTTTGTTAGGATATATAAAGTAAAGCCTGTATACAGGAGAC 686
 QY 780 rLeuAspHisLysProArgValThrAsnIlePhePro 792
 Db 687 ATTAGATCAAAACCTCGAGTCACCAACATTTTCCC 723

RESULT 11
 CF893672
 LOCUS 645 bp mRNA linear EST 04-NOV-2003
 DEFINITION A0126C03-5 NIA Mouse Undifferentiated ES Cell cDNA Library (Long 1)
 Mus musculus cDNA clone NIA:A0126C03 IMAGE:30726074 5', mRNA
 sequence.
 ACCSSION CF893672.1 GI:38160721
 VERSION CF893672
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 1 (bases 1 to 645)
 Construction of long-transcript enriched cDNA libraries from
 submicrogram amounts of total RNAs by a universal PCR amplification
 method
 Genome Res. 11 (9), 1553-1558 (2001)
 MEDLINE 21429098
 PUBMED 11544199
 COMMENT Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: A0126 row: C column: 03
 Seq primer: M13 Reverse
 High quality sequence stop: 645
 POLYA=No.

FEATURES Location/Qualifiers
 1. 645
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="129/Sv x 129/Sv-CP"
 /db_xref="niaEST:A0126C03-5"
 /db_xref="taxon:10090"
 /clone="NIA:A0126C03 IMAGE:30726074"
 /dev_stage="R1 ES cells"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse Undifferentiated ES Cell cDNA
 Library (Long 1)"
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
 Site 2: NotI; Mouse cDNA project by the Laboratory of
 Genetics, National Institute on Aging (NIA), Intramural
 Research Program, NIH (<http://igsun.grc.nia.nih.gov/cDNA>).
 This is a long-transcript enriched cDNA library (Ref.
 Genome Res. 11, 1553-1558 (2001). [PMID: 11544199]). Total
 RNAs were obtained from Dr. Kenneth R. Boheler (National
 Institute on Aging, USA). ES cells were cultured without

QY 394 sileHisleProtleleleAlaserValserGluHisGlnProThrThrTrpValSerPh 414
 Db 541 AATACACATTCATTTATTCATCAGTGTCTGACATCAGCCACGACATGGGTCTTT 600
 QY 414 ePhePheAspLeuHisleleValCysThrPheProAlaGlyLeuTrpPheCyslele 434
 Db 601 CTTCTTTGATCTACATATCTGTGTATGACTTCCCGACGAGGCTATGGTCTGCATCA 660
 QY 434 sAnileAsnAspGluArgValPheValAlaLeuTyAlaIleSerAlaValTyPheAl 454
 Db 661 AAATATCAACGATGAAGAGGTTTGTGCTGTGTATGCGATCAGTGCCTACATTGC 720
 QY 454 a 454
 Db 721 C 721

RESULT 9
 LOCUS CAS11417 658 bp mRNA linear EST 15-NOV-2002
 DEFINITION UI-R-FJ0-cpw-p-10-0-UI.r1 UI-R-FJ0 Rattus norvegicus cDNA clone
 CAS11417
 ACCSSION UI-R-FJ0-cpw-p-10-0-UI 5', mRNA sequence.

VERSION CAS11417.1 GI:25002371

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 658)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL discovery

MEDLINE Genome Res. 6 (9), 791-806 (1996)

PUBMED 9704477

COMMENT 8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: Bento-soares@uiowa.edu

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 REVERSE.

Location/Qualifiers

1..658

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-FJ0-cpw-p-10-0-UI"

/tissue_type="embryo"

/dev_stage="embryo"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-R-FJ0"

/note="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;

UI-R-FJ0 is a cDNA library containing the following

tissue(s): rat embryo. The library was constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed,

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into p77T3-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CATCTCTACT. This library
 was created for the University of Iowa Program for Rat
 Gene Discovery and Mapping (Val Sheffield, Bento Soares
 and Tom Casavant)"

ORIGIN

Alignment Scores:

Pred. No.: 2,04e-225 Length: 658
 Score: 219.00 Matches: 219
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 26.51% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x CA5111417 (1-658)

QY 218 GlySerPheAspAsnGluGlyLeAlaIlePheAlaLeuGlnPheThrTyTrpLeuTrp 237
 Db 2 GGATCCCTTTGATACGAGGCAATTCGCACTTCAGTTCACTTCACTTCACTTATGG 61
 QY 238 VallysSerVallyThrGlySerValPheThrMetCysCysLeuSerTyPhe 257
 Db 62 GTGAAGTCTGTGAAACCGGGTCTGTGTTGGCAATGCTGCTGCTGCTGCTATTC 121
 QY 258 TyrMetValSerAlaTrpGlyGlyTyValPheIleIleLeuLeuProLeuHisVal 277
 Db 122 TACATGCTCTGCGTGGGAGGTTATGTTTCATCATCAACCTCATCCCTCCATGTG 181
 QY 278 PheValLeuLeuMetGlnAtgTySerIyArgValTyIleAlaTySerThrPhe 297
 Db 182 TTTGCTGCTCTCCGATGACAGAGTACGCAAGAGCTTACATAGATAGACATTC 241
 QY 298 TyrlleValGlyLeuLeuLeuSerMetGlnIleProPheValGlyPheGlnPheArg 317
 Db 242 TACATTGCGGTTTATATATATCCATCAGATACCTTTTGTGGATTCCAGCCATCAGG 301
 QY 318 ThrSerGluHisMetAlaAlaGlyValPheAlaLeuGlnAlaTyAlaPheLeu 337
 Db 302 ACAAGCAGACATGGCAGCTGCGAGGTGCTTTGCGCTGCTGAGGCTTATGCGTTCTG 361
 QY 338 GlnTyLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPheLeuGlyVal 357
 Db 362 CAGTATCTGAGAGACCGTTTGACAAAGCAGAGTTCAGACACCTTTCTTTTGGGTGTC 421
 QY 358 SerleuAlaGlyAlaValPheLeuSerValIleTyLeuThrTyTrpGlyTyrlle 377
 Db 422 TCCTAGCTCAGCGCGTGTGTTCTTCTAGTGTCTATCTGACATACACAGGTACATT 481
 QY 378 AlaProTrpSerGlyArgPheTySerLeuTrpAspThrGlyTyAlaIleHisIle 397
 Db 482 GCACCATGGAGTGGCAGGTTTATTATTATGGGATCTGGGTATGCAAAATACACATT 541
 QY 398 ProIleIleAlaSerValSerGluHisGlnProThrTrpValSerPhePheAsp 417
 Db 542 CCAATTATTGTCATCAGTGTCTGAACATCAGCTTCAGACATGGGTGCTTTCTTTTGAT 601
 QY 418 LeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleIleHisIle 436
 Db 602 CTACATATTCTTGTGTGTTACCTTCCCGAGGCGCTATGTTTCTGCATCAAAATATC 658

RESULT 10

CF995035

LOCUS

DEFINITION

IMAGS:30528919 5', mRNA sequence.

CF995035

VERSION

CF995035.1 GI:38511095

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 861)

CF995035 861 bp mRNA linear EST 25-NOV-2003

IMAGS:30528919 5', mRNA sequence.

CF995035

VERSION

CF995035.1 GI:38511095

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 861)

us-10-028-384-2.Oligo.rst

Thu Dec 16 16:25:06 2004

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 DNA Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Alignment Scores: 2,748-237 Length: 693
 Pred. No.: 230.00 Matches: 230
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 27.85% Indels: 0
 Query Match: 7 Gaps: 0
 DB:

US-10-028-384-2 (1-826) x CN461211 (1-693)

QY 163 ThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSer 182
 DB 3 ACAGTTTACATAAGAGATGTGTGTATTCCTTGCACCAACTTTTATGCGGCTTACATCC 62
 QY 183 IleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAla 202
 DB 63 ATATCTACCTTCTGCTAATAGAGAACTGTGGAAACCAAGGAGGAGGACTTCTAGTCT 122
 QY 203 CysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsn 222
 DB 123 TCGTTCATTGCTATCTACAGGACATATCTCGTCACTGCGGAGTCTTTCATTAAT 182
 QY 223 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpVallySerVally 242
 DB 183 GAAGGCAATTCGCCATTTTGGCTTCAGTTTCACTTACTTATGGGTAAAGTCTGTGAAG 242
 QY 243 ThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
 DB 243 ACCGGTCTGTGTTCTGGCAATGTCTGCTGCTGTATATTTCTACATGCTCTGCG 302
 QY 263 TrpGlyGlyTyrValPheIleLeuLeuLeuLeuProLeuHisValPheValLeuLeu 282
 DB 303 TGGGAGGTTATGTGTTCATCATCAACCTCATCCCTCTCCATGTGTGTGTGTGTGTG 362
 QY 283 MetGlnArgTyrSerIleValGlyValTyrIleAlaTyrSerThrPheTyrIleValGly 302
 DB 363 ATGAGAGGTACAGAGAGAGAGTCTACATAGCATATAGACATTTCTACATGCTGGT 422
 QY 303 IleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 322
 DB 423 ATATATCCATGTCAGATACCTTTTGTGGGATTTTCAGCAATCAGAACAGGAGCAG 482
 QY 323 AlaAlaIleValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAsp 342
 DB 483 GCAGCTCGAGGTGTCTTGGCGTCTGCAAGCTTACGCTTTTGTGAGTATCTGAGAG 542
 QY 343 ArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGly 362
 DB 543 CGGTTGACAAAACAGGAGTTCAGACCTTTTCTTTTGGGTGTCTCACTAGTGCAGGC 602
 QY 363 AlaValPheLeuSerValIleTyrLeuThrTyrTyrGlyTyrIleAlaProThrPheSerGly 382
 DB 603 GCTGTGTTCTTCTAGTGTATCTATCTGACATACAGGTTATATGACCATGGAGTGC 662
 QY 383 ArgPheTyrSerLeuTrpAspThrGlyTyr 392
 DB 663 AGGTTTATTCATCTGGGATCTGGGTAT 692

RESULT 8
 BU515954
 LOCUS AGNCCOURT 10120433 NIH MGC 134 Mus musculus cDNA clone
 DEFINITION IMAGE:6512642 5', mRNA sequence.
 BU515954
 ACCESSION BU515954.1 GI:22823480
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE 1 (bases 1 to 950)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source

1..950
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6512642"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_134"
 /notes="Vector: pCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dT. Average insert
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
 this is a NIH_MGC Library."

ORIGIN

Alignment Scores: 6,648-234 Length: 950
 Pred. No.: 227.00 Matches: 240
 Score: 99.59% Conservative: 0
 Percent Similarity: 99.59% Mismatches: 0
 Best Local Similarity: 27.48% Indels: 1
 Query Match: 5 Gaps: 0
 DB:

US-10-028-384-2 (1-826) x BU515954 (1-950)

QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
 DB 1 TCAGTGGCGGAGTCCCTTGTATATGATGAGGATTCGCCATTTTCGGCTTCAGTTCATTAC 60
 QY 235 TyrLeuTrpVallySerVallyThrGlySerValPheTrpThrMetCysCysLeu 254
 DB 61 TACTTATGGGTAAAGTCTGTGAAGACCGGCTCTGTCTCGCAATGTGCTGCTGCTTG 120
 QY 255 SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleLeuLeuLeuPro 274
 DB 121 TCATATTTCATATGTTCTCTGGTGGGAGGTTATGTGTTTCATCATCAACCTCATCCCT 180
 QY 275 LeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
 DB 191 CTCCATGTGTTTGTGTTGCTGCTGATGACAGGTACAGCAAGAGAGTCTACATAGCATAT 240
 QY 295 SerThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGln 314
 DB 241 AGCATTTTCTACATGTGGGTTTAATATATCCATGCAATACCTTTTGTGGGATTTTTCAG 300
 QY 315 ProIleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuGlnAlaTyr 334
 DB 301 CCATCAACAACAGGAGCACATGCGAGTGTCTTTGGCTGTCTGCAAGCTTAC 360
 QY 335 Ala-PheLeuGlnTyrLeuArgAspArgLeuThrIleGlnGluPheGlnThrLeuPhePh 354
 DB 361 GCTTTTTTGCAGTATCTGAGAGACCGGTGACAAAAACAGAGTTCAGACCCCTTTTCTT 420
 QY 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyr 374
 DB 421 TTTGGGTGCTCTACATAGTGTGAGGCGGTGTCTTCTTAGTGTCTATCTATCTGACATAC 480
 QY 374 cGlyTyrIleAlaProThrPheSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLy 394
 DB 481 AGGTTATATGACCATGAGTGGCAGGTTTATTTACTATGGGATATCTGGGTATGCAAA 540

Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaops-remail.nih.gov
Tissue Procurement: Naryan Bhat
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM1098 row: d column: 22
High quality sequence stop: 710.

FEATURES

Location/Qualifiers
1..775
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30786813"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_222"
/note="Organ: placenta; Vector: pExpress-1; Site 1: EcoRV;
Site 2: NotI; RNA obtained from three placentas from
female C57/BL6 mouse at 16 days pregnancy. Tissues were
snap-frozen and kept at -80C for two days before RNA
extraction and purification (Tri-reagent method). cDNA was
primed using oligo-dT primer:
5'-TGACTAGTTCATGTCGCGCGCGCCG(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. Size-selection >1 kb
resulted in an average insert size of 1.5 kb. Library is
not amplified. (Normalized version of this library is
NIH MGC 203.) Library constructed by Express Genomics
(Frederick, MD)."

ORIGIN

Alignment Scores:
Pred. No.: 2,11e-239 Length: 775
Score: 232.00 Matches: 232
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 28.09% Indels: 0
DBs: 7 Gaps: 0
US-10-028-384-2 (1-826) x CK032376 (1-775)

Qy 194 AsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSer 213
Db 2 AACCAAGGAGGAGGACTTCTAGCTGCTTCTTCTATGCTATGCTACGAGGTACATATCT 61
Qy 214 ArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 233
Db 62 CGGTCACTGGCGGATCTTGTATGATGAAGGCAATGCCATTTTGGCTTCAGTTCACT 121
Qy 234 TyrTyrLeuTrpValLysSerValLysThrGlySerValPheThrMetCysCysCys 253
Db 122 TACTACTTATGGTAAAGTCTGTGAAGCGGGTCTGTGTCTGGCAATGTGCTCTGTC 181
Qy 254 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleLeuIle 273
Db 182 TTGTCAATTTCTACATGGTCTTCTGGTGGGAGGTATGTTCTATCATCACTCATC 241
Qy 274 ProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAla 293
Db 242 CCTCTCCATGTGTTTGTGTTCTGCTATGTCAGAGGTACAGCAAGAGGTCTACATAGCA 301
Qy 294 TyrSerThrPheTyrIleValGlyLeuLeuLeuSerMetGlnIleProPheValGlyPhe 313
Db 302 TATACACTTCTACATGTGGGTTTATATATCAATGAGATACCTTTTGGGATTT 361
Qy 314 GlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAla 333
Db 362 CAGCCCAATCAGAAAGGAGGACATGCGAGCTGAGGTCTTCTTGGCTGTGCTGAAGCT 421

Qy 334 TyAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhe 353
Db 422 TAGGCTTTTTCAGGTATCTGAGAGCCGGTTGACAAACAGGAGTTCACAGCCCTTTTC 481
Qy 354 PheLeuGlyValSerLeuAlaIleAlaValPheLeuSerValIleTyrIleThrTyr 373
Db 482 TTTTGGGTTCTCACTAGCTGAGCGGCTGTGTTCCCTTAGTGTCTATCTGACATAC 541
Qy 374 ThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 393
Db 542 ACAGGTTATATGACCATGGAGTGGCAGGTTTATTCTACATATGGGATACCTGGGTATGCA 601
Qy 394 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrValSer 413
Db 602 AAAATACATATCCAAATTATTCATCAGTGTCTGACATCAGCTACCATGGGTGCT 661
Qy 414 PhePhePheAspLeuHisIleLeuValCysThrPhe 425
Db 662 TTCTTCTTGTATCATATATCTTGTATGATACCTTC 697

RESULT 7

CN461211 693 bp mRNA linear EST 21-APR-2004
UI-M-HNO-cot-g-05-0-UI.r1 NIH_BMAP_HNO Mus musculus cDNA clone
IMAGE:30652468 5', mRNA sequence.
CN461211
CN461211.1 GI:46466937
EST.
Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 693)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaops-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP).

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1..693
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30652468"
/tissue_type="Upper Head"
/dev_stage="9.5-10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HNO"
/note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGACATGAT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN

374 ThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 393
 543 ACAGGTTATATTCACCATGGAGTGGCAGGTTTATTCACATGGGATCTGGGTATGCA 602
 394 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpValSer 413
 603 AAATACACATTCCTCAATATTCATCAGTGTCTGAACATCAGCCATCCACATGGGTGCT 662
 414 PhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIle 433
 663 TTCTTCTTGAICTACATATTTCTGTATGATCTTCCAGCAGGCTATGGTTCGATC 722

BE952197 747 bp mRNA linear EST 29-APR-2002
 LOCUS UI-M-CDO-aj1-h-06-0-UI.81 NIH BMAP Ret2 Mus musculus cDNA clone
 DEFINITION UI-M-CDO-aj1-h-06-0-UI 3', mRNA sequence.

ACCESSION BE952197

VERSION BE952197.1 GI:10592357

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 747)

TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.

JOURNAL Normalization and subtraction: two approaches to facilitate gene

MEDLINE Genome Res. 6 (9), 791-806 (1996)

PUBMED 97044477

COMMENT 889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mESToma1.nih.gov

Oligo-dT track not found. Not I site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
 clones from RESEARCH GENETICS. It should be noted that Bento Soares
 is generating a small number of additional specialized
 non-redundant arrays of BMAP cDNAs whose availability will be
 considered under appropriate and limited collaborative arrangements
 The tissue for this library was contributed by Dr. Xin-Yuan Fu,
 Yale University School of Medicine
 Seq primer: M13 Forward
 POLYA=No.

FEATURES

source

1. 747
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-CDO-aj1-h-06-0-UI"
 /dev_stage="1 day"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NIH BMAP Ret2"
 /note="Vector: pT7m3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The
 NIH BMAP Ret2 library is derived from mouse neonatal
 retina tissue. For a detailed description of the library
 from which this clone was derived, please visit our web
 site at brainest.eng.uiowa.edu. The tissue for this
 library was contributed by Dr. Xin-Yuan Fu, Yale
 University School of Medicine
 TAG_SEQ=None found"

ORIGIN

Alignment Scores:

Pred. No.:

1.16e-242

Length:

747

Score: 235.00 Matches: 235
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 28.45% Indels: 0
 DB: 2 Gaps: 0

US-10-028-384-2 (1-826) x BE952197 (1-747)

Qy 157 LeuAsnThrLeuAsnIleThrValHisIleAArgAspValCysValPheLeuAlaProThr 176

Db 24 TTAATACATTTGAACATAACAGTTACATAAGAGATGTGTGTATTCCTTGACCAACT 83

Qy 177 PheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGly 196

Db 84 TTTAGCGGCTTACATCCATATCTACGTTCTCTTAAGTCTAGAACTGTGGAACCAAGGA 143

Qy 197 AlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerVal 216

Db 144 GCAGGACTTCTAGCTGCTTCTTCATTGCTATCGTACCAGGGTACATATCTCGGTGAGTG 203

Qy 217 AlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrIleLeu 236

Db 204 CGGGATCCCTTGAATATGAGSCATGTCATTTCGCTTCAGTTCACCTACTTACTTA 263

Qy 237 TrpValIysSerValIysThrGlySerValPheTrpThrMetCysCysCysSerIleTyr 256

Db 264 TGGTAAAGTCTGTGAAGACCGGCTGTGTTCTTGACCAATGTGCTGCTGCTTGCATAT 323

Qy 257 PheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHis 276

Db 324 TTCTACATGCTCTCGGTGGGAGGTTATGTTTCATCATCAACCTCATCCCTCTCCAT 383

Qy 277 ValPheValLeuLeuLeuMetGlnAArgTyrSerLysArgValTyrIleAlaTyrSerThr 296

Db 384 GTGTTTGTGTTGCTGTGATGCAGAGGTACAGACAGAGTCTACATAGCATATAGCACT 443

Qy 297 PheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIle 316

Db 444 TTCTACATTTGGGTTTAAATATTCATCCATGCAGATACCTTTTGGGATTCAGCCAAATC 503

Qy 317 ArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPhe 336

Db 504 AGAACAGCGAGCACATGGCAGTGCAGGTTCTTTGGCTGCTGCAAGCTTACGCTTT 563

Qy 337 LeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGly 356

Db 564 TTGCAGTATCTGAGAGACCGGTTGACAAAACAGAGTTCAGACCCCTTTTCITTTGGGT 623

Qy 357 ValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyr 376

Db 624 GTCTCATAGCTGCAGCGGCTGTGTTCTTATGTTGTCATCTATCTGACATACACAGTTAT 693

Qy 377 IleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGly 391

Db 684 ATTGCACATGGAGTGGCAGGTTTATTCATCTACTATGGGATACTGGG 728

RESULT 6

CK032376

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CK032376 775 bp mRNA linear EST 26-NOV-2003
 AGENCOURT 16650470 NIH MGC 222 Mus musculus cDNA clone
 IMAGE:30786813 5', mRNA sequence.

CK032376

CK032376.1 GI:38558300

EST.

Mus musculus (house mouse)

Mus musculus

Mus musculus

1 (bases 1 to 775)

NIH-MGC <http://mgs.nci.nih.gov/>

National Institutes of Health

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

```

Db 1583 CACTGCACGTGGGTCACAAAGCGCTACTCCAGTCCAAAGTGGTCTTGCCTCTAC 1742
Qy 573 AenHisAspGlyThrArgAsnHisLeuAspAspPheArgGluAlaTyrPheTrpLeuArg 592
Db 1743 AATCATGATGTTACAGGAATATATTAGATGATTTAGAGAACGCTACTTTGGCTGAGA 1802
Qy 593 GlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnHisAla 612
Db 1803 CAACAACACGGATGAACACGCCGGCTCATGCTGGTGGGACTACGCGTATCATGATTGCT 1862
Qy 613 GlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisLeuAla 632
Db 1863 GGCAATGCCAACAGACCACTCTGGTGGATGAACACCTGGAGAACACACCCATCGCA 1922
Qy 633 LeuValGlyAlaMetSerSerAsnGluThrAlaAlaTyrLysLeuMetArgThr-Le 652
Db 1923 CTGGTCGGAAGAAGCTATGCTTCCATGAACGCCGCCCTATAAATCATGAGTC-CCT 1981
Qy 652 uAspValAspTyrValLeuValLlePheGlyGlyValLleGlyTyrSerGlyAspAsp 1981
Db 1982 TGATGTCGATTAATGCTGTTGTTATTTCCGAGGAGTATGGCTATTCGGGGACGATAT 2041
Qy 672 eAsnLysPheLeuTrpMetValArgLeuAlaGluGlyGluHisProLysAspLeuArg 692
Db 2042 CAACAAGTTCCTCTGGATGTCAGGATAGCTGAAGGGGAGCATCCCAAGACATCCGGA 2101
Qy 692 uSer-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThr 711
Db 2102 AGG-TGACTATTTTAC-CGACGAGGAGAGTTCGAGTAGACAAAGCTGGTCTCTACT 2159
Qy 712 LeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
Db 2160 CTGTAAACGCTTATGATATAAATGTCATACATCAGATTGGAGAAATGACGTAGAT 2219
Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluLeuGlyAsnLysAsp 751
Db 2220 TTTCGCACCTCCCGAGCTTTGACCAACACGCTAATGCTGAGATGGAAATAAGACATT 2279
Qy 752 LysPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgLysTrpLys 771
Db 2280 AAATTCAAGCATTTGAGGAGAGCTTTTACATCAGACATGCTGCTGTCAGGATATATA 2339
Qy 772 ValLysAlaProAspAsnArgGluThrLeu 781
Db 2340 GTGAAAGCCTGACAAACAGGGAGACACTA 2369

```

```

RESULT 4
CN458732
LOCUS
DEFINITION
  CN458732 852 bp mRNA linear EST 21-APR-2004
  UI-M-HBO-cov-1-02-0-UI.r1 NIH_BMAP_HBO Mus musculus cDNA clone
  IMAGE:30649513 5', mRNA sequence.
ACCESSION
  CN458732
VERSION
  CN458732.1 GI:46464458
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
  Mus musculus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 bases 1 to 852
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaaps-r@mail.nih.gov
  Tissue Procurement: Dr. James Lin University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Distribution information can be found at
  http://genome.uiowa.edu/distribution/mousefl.html
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)

```

FEATURES

```

Seq primer: pYX-5.
Location/Qualifiers
  1..852
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="CS7BL/6"
  /db_xref="taxon:10090"
  /clone="IMAGE:30649513"
  /tissue_type="whole eye"
  /dev_stage="embryo 12.5, 13.5, 14.5 dpc"
  /lab_host="DH10B (T1 phage resistant)"
  /clone_lib="NIH_BMAP_HBO"
  /note="Organ: Eye; Vector: pYX-Asc; Site 1: Ecor I;
  Site 2: Not I; The library was constructed according
  Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,
  1996. Denatured RNA was size fractionated on a 1% agarose
  gel. First strand cDNA synthesis was primed with oligo-dT
  primer containing a Not I site. Double strand cDNA was
  size selected according to mRNA size fraction, ligated
  with Ecor I adaptor, digested with NotI and then cloned
  directionally into pYX-Asc vector. The library tag
  sequence located between the Not I site and the polyA tail
  is TTAATTGAAGT. This library was created for the University
  Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
  Developing Mouse Nervous System', supported by National
  Institute of Mental Health (NIMH)."
```

ORIGIN

```

Alignment Scores:
Pred. No.: 5,19e-248 Length: 852
Score: 240.00 Matches: 240
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.06% Indels: 0
Db: 7 Gaps: 0
US-10-028-384-2 (1-826) x CN458732 (1-852)

```

```

Qy 194 AsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSer 213
Db 3 AACCAAGAGCAGGACTTCCTAGCTGCTTGTCTTCAATGCTATCTGACAGGTCATATCT 62
Qy 214 ArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 233
Db 63 CGGTCACTGGCGGATCTCTTGTATATGAAGGATGTCATTTTGGCTTCAGTTCAC 122
Qy 234 TyrTrpLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCysCys 253
Db 123 TACTACTATGGGTAAAGTCTGTGAAGACCGGCTCTGTGTCTGGACAATGTGCTGTC 182
Qy 254 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuLe 273
Db 183 TTGTCATATTTTACATGGTCTCTGGTGGGAGGTTATGTGTTCATCATCAACCTCATC 242
Qy 274 ProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAla 293
Db 243 CCTCTCCATGTTTGTGTGCTGCTGATGACAGGTCAGAGAGAGAGTACATAGCA 302
Qy 294 TyrSerThrPheTyrIleValGlyLeuLeuLeuSerMetGlnIleProPheValGlyPhe 313
Db 303 TATAGCACTTTTCTACATGTTGGGTTTAAATATTCATCATGACATACCTTTTGGGAT 362
Qy 314 GlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAla 333
Db 363 CAGCCAATCAGAAACAGCAGGACATGGCAGCTGCAGGTTGCTTTTGGCTGTCGCAAGCT 422
Qy 334 TyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhe 353
Db 423 TAGCTTTTTCAGTATCTGAGAGACCGGTGACAAAACAGAGGTTCCAGACCCCTTTTC 482
Qy 354 PheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyr 373
Db 483 TTTTGGGTGCTCTCACTAGCTGACGCGCTGTGTTCTTCTAGTGTCTATCTGACATAC 542

```


ORIGIN

Alignment Scores:
 Pred. No.: 5,24e-293 Length: 2212
 Score: 282.00 Matches: 282
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 34.14% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-2 (1-826) x CR610915 (1-2212)

QY 545 MetLeuMetMetPheAlaValHisCysThrTrpValThrSerAlaTyrSerSer 564
 DB 3 ATGCTATTGATGATGTTTGGCTGCTCAGTACCTGGGTCAAGCAATGCTACTTAGT 62
 QY 565 ProSerValLeuAlaSerTyrAsnHisAspGlyThrArgAsnLeuLeuAspPhe 584
 DB 63 CCAAGTGTAGTCTGGCTCCTCATAAATCATGATGGCCAGCAATATCTTAGATGATTT 122
 QY 585 ATGGGAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrp 604
 DB 123 AGAGAGGCTTACTTTGGCTAGGCAAAATACAGATGACATGCACAGATTAATGCTTGG 182
 QY 605 TrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsn 624
 DB 183 TGGGATTATGGCTATCATAGTCTGGAATGGCTAATAGAACTACGTTGGTGGATAATAAC 242
 QY 625 ThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAlaGluThrAla 644
 DB 243 ACTTGGATTAACAGCCATATAGCACTGGTGGGAAGCAATGCTTCTTAATGAAACGCA 302
 QY 645 AlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyVal 664
 DB 303 GCGTATAAAATCATGAGCACTAGATGATGATTTGTTGGTATTTTGGAGGGGTT 362
 QY 665 IleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGly 684
 DB 363 ATGGCTATTCTGGTGAATGATATCAAAATTTCTTGGATGGTTAGTAGTGAAGGA 422
 QY 685 GluHisProLysAspIleArgLysSerAspTyrPheThrProGlnGlyLysPheArgVal 704
 DB 423 GAACATCCCAAGACATTCGGGAAGTCACTATTATCCCAAGAGGAGAAATTCCTGTGA 482
 QY 705 AspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTrpArg 724
 DB 483 GACAAAGCAGGATCCCTTACTTTGTTGATTCCTTATGATATAAATGCTCATCTACAGA 542
 QY 725 PheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAla 744
 DB 543 TTTGGAGAAATGCAGCTGGATTTTGGTACACCCCGAGGTTTGGACCGAACACGTAATGCT 602
 QY 745 GluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHis 764
 DB 603 GAGATGGAAATAGGACATTAATTCAAATTCATTTGGAAGAGCCCTTATCATCAGAACAC 662
 QY 765 TrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLys 784
 DB 663 TGGCTTTTAGGATATATAAAGTAAAGCACCTGATTAACAGGGAGACATTAGATCAAAA 722
 QY 785 ProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArg 804
 DB 723 CTCGAGTCACCAACATTTTCCCAACAGAGATTTTGTCAAGAGAGACTACCAAGAGG 782
 QY 805 LysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLys 824
 DB 783 AAGCGTGGCTACATATAAATAAGTGGTGTGTTTAAAGAAAGCAAGAAATATCTAAGAG 842
 QY 825 ThrVal 826
 DB 843 ACTGTT 848

RESULT 3

AK012153

LOCUS

DEFINITION

AK012153 2669 bp mRNA linear HTC 03-APR-2004
 Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
 enriched library, clone:2610524N02 product:hypothetical
 Oligosaccharyl transferase (OTase) STT3 subunit containing protein,
 full insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK012153.1 GI:12848721

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

COMMENT

The RIKEN Genome Exploration Research Group Phase II Team and the
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases:1 to 2669)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
 Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
 Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
 Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
 Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
 Terajima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
 Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.jp,
 URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

Please visit our web site (<http://genome.gsc.riken.jp/>) for further

QY 255 SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIlePro 274
 DB 829 TCATATTTCTACATGGTCTCTGGCTGGGAGGTTATGTTTCATCATCAACCTCATCCCT 888
 QY 275 LeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
 DB 889 CTCCTATGTTGTGTGTCTCTGATGAGAGGTACAGCAAGAGAGTCTACATAGCATAT 948
 QY 295 SerThrPhe-TyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheG 314
 DB 949 AGCAC-TTTGTACATTTGGGTTTATATATATCCATGCAGATACCTTTTGGGATTCA 1007
 QY 314 nProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyr 334
 DB 1008 GCCAATCAGAAACAGCAGACATCGCAGCTGCAGGTGTCTTTGCGCTGCTGCAAGCTTA 1067
 QY 334 rAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnPheGlnThrLeuPhePh 354
 DB 1068 CGCTTTTTCAGTATCTGAGACCGGTTGACAAACAGAGTTCCAGACCCCTTTTCT 1127
 QY 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrTh 374
 DB 1128 TTTGGGTGTCTCACTAGCTGAGCGCTGTCTTCTTGTAGTGTCTATCTATCTGACATAC 1187
 QY 374 rGlyTyrIleAlaProTtpSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAlaL 394
 DB 1188 AGGTATATTCACCATGGAGTGCAGGCTTTTATTCATATGGATACCTGGGTATGCATA 1247
 QY 394 sIleHisIleProIleAlaSerValSerGluHisGlnProThrTyrTrpValSerPh 414
 DB 1248 AATACACATTCATATTTGATCATGTGTGTGTAACATCAGCTACGACCTACGACGTGTCTTT 1307
 QY 414 ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTyrPheCysIleL 434
 DB 1308 CTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCTATGTTCTGCATCAA 1367
 QY 434 sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl 454
 DB 1368 AAATATCAACGATGAAAGAGTATTTGCTGCTGTATGCGATCAGTGTGTGTACTTTGC 1427
 QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaI 474
 DB 1428 CGGAGTGTGTGTGGGTGATGCTGATCTGACCCCGTGTCTGTGATGCTGTGCGCAT 1487
 QY 474 eAlaPheSerAsnValPheGluHisTyrLeuGlyAspMetLysArgGluAsnProPr 494
 DB 1488 CGCCTTCTCCATGTTTGTAGCACTATTTGGGGATGACATGAAAGGAGAAACCCACC 1547
 QY 494 oValGluAspSerSerAspGluAspLysArgAsnGln-GlyAsnLeuTyrAspLysA 514
 DB 1548 TGTGAGGACAGCAGTGTGATGAGGATGACAAAGAAACCC-AGGAACTTGTATGACAGG 1606
 QY 514 laGlyLysValArgGlyHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr 533
 DB 1607 CAGGTAAAGTGGAGGACATGT-GACACAGCAAGAGAA-ACCTCAAGAGGCTTGGGCC 1664
 QY 533 oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHi 553
 DB 1665 CAACATCAAAAGCATTTGACCATCTGTATGCTGTCTCTCTGATGATGTTTGGGTCCA 1724
 QY 553 sCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAs 573
 DB 1725 CTGCAGCTGGGTCAAGACACGCTACTCCAGTCCAAAGTGTGGTCTGCTTGCCTCTACAA 1784
 QY 573 nHisAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgG 593
 DB 1785 TCATGATGTACAGGATATATTAGATGATTTTAGAGAGCGTACTTTTGGCTGAGACA 1844
 QY 593 rAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaG 613
 DB 1845 AAACACGATGACACCGCCGGTGTGTGTGTGGTGTGGTGTGTGTGTGTGTGTGTGTGTGT 1904
 QY 613 yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLe 633

DB 1905 CATGCCAACAGGACCACTCTGTGTGATTAACAACCTTGAACAACAGCCATCGCACT 1964
 QY 633 uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuA 653
 DB 1965 GGTGGGAAAGCTATCTCTTCAATGAAACGGCCGCTATAAAATCATGAGTCC-CCITG 2023
 QY 653 spValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleA 673
 DB 2024 ATGTCGATTATGTTGGTTATTTTCGAGGAGTGTGATTTGCTTATTCGCGGACGATATCA 2083
 QY 673 snLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluS 693
 DB 2084 ACAAGTTCTCTCGATGGTTCAGATAGTGAAGGGGAGCATCCCAAGACATCCGGGAAG 2143
 QY 693 er-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe 712
 DB 2144 G-TGACTATTTTAC-CCAGCAGGAGAGTTCCGAGTAGACAAAGCTGGGTCTCTACTCT 2201
 QY 712 uLeuAsnCysLeuMetTyrIleMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPh 732
 DB 2202 GTTAAACTGCTTATGTATATAAATGTCTACTACAGATTTGGAGAAATGCAGTAGATT 2261
 QY 732 eaArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleL 752
 DB 2262 TCGACTCCCCCAGGCTTTGACCGAACAGTATCTGAGATTGGAATTAAGACATTAA 2321
 QY 752 sPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysVa 772
 DB 2322 ATTCAAGCATTTGGAGAAAGCTTTTACATCAGACAGCTGGCTGTGCAGGATATAAAGT 2381
 QY 772 lLysAlaProAspAsnArgGluThrLeu 781
 DB 2382 GAAAGCAGCTGACAAACAGGAGACACTA 2409
 RESULT 2
 CR610915 2212 bp mRNA linear HTC 21-JUL-2004
 LOCUS full-length cDNA clone CSODK010Y001 of HeLa cells Cot 25-normalized
 DEFINITION of Homo sapiens (human).
 ACCESSION CR610915
 VERSION CR610915.1 GI:50491722
 KEYWORDS HTC; CNSIT_cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 2212)
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact : feng liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue
 REFERENCE 2 (bases 1 to 2212)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 BP 131 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo (dr) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 source 1. 2212
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CSODK010Y001"
 /issue_type="HeLa cells Cot 25-normalized"
 /plasmid="pCMVSPORT_6"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:58 ; Search time 6394.05 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-2
Perfect score: 826
Sequence: 1 MAEPAPSSKHKSLNSLSPW.....GVIKNLVFKGKIKKTKV 826

Scoring table:

| | |
|----------------------------|----------------------------|
| OLIGO | Xgapop 60.0 , Xgapext 60.0 |
| Ygapop 60.0 , Ygapext 60.0 | |
| Fgapop 6.0 , Fgapext 7.0 | |
| Delop 6.0 , Delext 7.0 | |

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USFTO.spool/US10028384/runat_14122004.131646_6807/app_query.fasta_1.3740
-DB=EST -QFMT=fastap -SUFFIX=oligo -L=LOOPEL=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.edi -LIST=45
-LOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=proto
-NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US10028384 @CNG 1 1 13960 @runat_14122004.131646_6807 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hc:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gss1:*
- 9: gb_gss2:*

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 365 | 44.2 | 2709 | AK018758 | AK018758 Mus muscu |
| 2 | 282 | 34.1 | 2212 | 3 | CR610915 full-len |
| 3 | 243 | 29.4 | 2669 | 3 | AK012153 Mus muscu |
| 4 | 240 | 29.1 | 852 | 7 | CN458732 UI-M-HBO- |
| 5 | 235 | 28.5 | 747 | 2 | BE952197 UI-M-CDO- |
| 6 | 232 | 28.1 | 775 | 7 | CK032376 AGENCOURT |
| 7 | 230 | 27.8 | 693 | 7 | CN461211 UI-M-HNO- |
| 8 | 227 | 27.5 | 950 | 5 | BU515954 AGENCOURT |
| 9 | 219 | 26.5 | 658 | 6 | CA511417 UI-R-FJO- |

Result 1

LOCUS AK018758

DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300006C19 product:hypothetical oligosaccharyl transferase (OTase) SRT3 subunit containing protein, full insert sequence.

ACCESSION AK018758

VERSION AK018758.2 GI:26384577

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1. Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253

2. 10349636

REFERENCE 3. Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishire, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sugahara, Y., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

ALIGNMENTS

2709 bp mRNA linear HTC 03-APR-2004

AK018758 Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300006C19 product:hypothetical oligosaccharyl transferase (OTase) SRT3 subunit containing protein, full insert sequence.

AK018758.2 GI:26384577

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1. Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253

2. 10349636

REFERENCE 3. Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishire, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sugahara, Y., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

Db 62 ATAGCTGGATGGCTAATAGAACTACGTTGGTGGATAAACAACCTGGATTAACAGCCAC 121
 Qy 631 IleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArg 650
 Db 122 ATAGCACTGGTGGGAAAAGCTATGCTCTTCTAATGAACAGCAGCCCTATAAAATCATGAGG 181
 Qy 651 ThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 670
 Db 182 ACTCTAGATGATAGATTATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGAT 241
 Qy 671 AspIleAsnLysPheLeuThrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
 Db 242 GATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAGAACATCCCAAGACATT 301
 Qy 691 ArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerPro 710
 Db 302 CGGGAAGTGACTATTTTACCCACAGGGGAGAAATTCGGTGTAGACAAAGCAGGATCCCT 361
 Qy 711 ThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeu 730
 Db 362 ACTTTGTTGAATGGCTTATGTATAAAATGTCATACAGATTTGGAGAAATGCAGCTG 421
 Qy 731 AspPheArg 733
 Db 422 GATTTTCGT 430

RESULT 15

US-10-046-935-717
 ; Sequence 717, Application US/10046935
 ; Publication No. US20020156011A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yuchi
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Stolk, John A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.527C1
 ; CURRENT APPLICATION NUMBER: US/10/046,935
 ; CURRENT FILING DATE: 2002-01-15
 ; NUMBER OF SEQ ID NOS: 2239
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 717
 ; LENGTH: 430
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-046-935-717

Alignment Scores:
 Pred. No.: 3,17e-141 Length: 430
 Score: 143.00 Matches: 143
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.31% Indels: 0
 DB: 13 Gaps: 0

US-10-028-384-2 (1-826) x US-10-046-935-717 (1-430)

Qy 591 LeuArgGlnAsnThrAspGluHisAlaArgValMetSerTyrTyrAspTyrGlyTyrGln 610
 Db 2 CTAAGGCAAAATACAGATGAACATGCACGAGTAATGCTCTTGGTGGGATTAATGGCTATCAG 61
 Qy 611 IleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTyrAsnAsnSerHis 630
 Db 62 ATAGCTGGATGGCTAATAGAACTACGTTGGTGGATAAACAACCTGGATTAACAGCCAC 121
 Qy 631 IleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArg 650
 Db 122 ATAGCACTGGTGGGAAAAGCTATGCTCTTCTAATGAACAGCAGCCCTATAAAATCATGAGG 181
 Qy 651 ThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 670

Db 182 ACTCTAGATGATAGATTATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGAT 241
 Qy 671 AspIleAsnLysPheLeuThrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
 Db 242 GATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAGAACATCCCAAGACATT 301
 Qy 691 ArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerPro 710
 Db 302 CGGGAAGTGACTATTTTACCCACAGGGGAGAAATTCGGTGTAGACAAAGCAGGATCCCT 361
 Qy 711 ThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeu 730
 Db 362 ACTTTGTTGAATGGCTTATGTATAAAATGTCATACAGATTTGGAGAAATGCAGCTG 421
 Qy 731 AspPheArg 733
 Db 422 GATTTTCGT 430

Search completed: December 15, 2004, 15:26:48
 Job time : 1072.95 secs

Qy 630 HisllealaleuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
 Db 423 CACATAGCACTGGTGGGAAAGCTATGCTCTTAATGAAACAGCAGCCTATAAAATCATG 482
 Qy 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
 Db 483 AGNACTCTAGATGAGTAAATGTTTGGTATTTTGGAGGGTATTTGGCTATTCGT 542
 Qy 670 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 689
 Db 543 GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 602
 Qy 690 IleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySer 709
 Db 603 ATTCCGGAAGTAGCTATTTTACCCACAGGGAGAAATTCGTTGTAGACAAAGCAGGATCC 662
 Qy 710 ProThrLeuLeu-AsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGly 729
 Db 663 CCTACTTTGKT-GAATTTGCCCTTATGTTAATAATGTCATCTACAGATTTGGAGAAATGCA 721
 Qy 729 nLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
 Db 722 GCTGGATTTCTGACACCCCGAGTTTGACCGAACACAGTAATGCTGAGATTTGGAAATAA 781
 Qy 749 sAepIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTyrLeuValArgIle 769
 Db 782 GGACATTAATTAACACATTTGGAAGAGCCCTTTACATCAGAACACTGGCTTTGTTAGGAT 841
 Qy 769 eTyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAs 789
 Db 842 ATATAAGTAAGACCACTGATTAACGGGAGACATTTAGATCACAACCTCGAGTCACCAA 901
 Qy 789 nilePheProLysGlnLysTyrLeuSerLysLysThrLysArgLysArgGlyTyrIle 809
 Db 902 CATTTTCCCAACAGAGATTTTGTCAAGAGAGACTACCAAAAGAAAGCGGTGGCTACAT 961
 Qy 809 eLysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThrVal 826
 Db 962 TAAAAAATAGCTGGTTTAAAGAGGCAAGAAATATCTAAGAGACCTGTT 1013

RESULT 13

US-09-998-598-1643
 ; Sequence 1643, Application US/09998598
 ; Patent No. US20020150922A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Chenault, Ruth A.
 ; APPLICANT: Meagher, Madelein Joy
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.561
 ; CURRENT APPLICATION NUMBER: US/09/998,598
 ; CURRENT FILING DATE: 2001-11-16
 ; NUMBER OF SEQ ID NOS: 2606
 ; SOFTWARE: Corixa Invention Disclosure Database
 ; SEQ ID NO 1643
 ; LENGTH: 500
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-998-598-1643

Alignment Scores:
 Pred. No.: 1,638-165 Length: 500
 Score: 166.00 Matches: 166
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.10% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x US-09-998-598-1643 (1-500)

Qy 158 AsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPhe 177
 Db 2 AATACATTGAACATAAATGTTTACATAAAGAGACGCTATGTTGTTCTTGCACCAACTTTT 61
 Qy 178 SerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsnGlnGlyAla 197
 Db 62 AGGGGCTTACATCATATATCTACTTTCTCTTACAGAGAACTTTGGAACCAAGGAGCA 121
 Qy 198 GlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAla 217
 Db 122 GGAATTTTACGCTGCTGTTTATTTCTATTGTACAGGCTACATCTCGTTCAGTAGCT 181
 Qy 218 GlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyr 237
 Db 182 GGATCTCTTGAATAAGAGGCATTCGTTATTTTGCACCTTCAGTTCACATCTATTATGG 241
 Qy 238 ValLysSerValLysThrGlySerValPheThrMetCysCysLeuSerTyrPhe 257
 Db 242 GTAAATCTGTAAAAACATGGGTCAGTTTTTTGGACAATGTCTGCTGCTTATCTATTTC 301
 Qy 258 TyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIleProLeuHisVal 277
 Db 302 TATATGGTCTCTGCTTGGGGTGGTTATGTTATTTATCATCAATCTTATTCACATGTA 361
 Qy 278 PheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPhe 297
 Db 362 TTTGTGTTGTTACTGATGCAGAGATACAGCAAAAGAGCTACATAGCATATAGCACTTTC 421
 Qy 298 TyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnProIleArg 317
 Db 422 TACATTTGGTGGTTTAAATTAATTAATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGA 481
 Qy 318 ThrSerGluHisMetAla 323
 Db 482 ACAAGTGAACACATGCGA 499

RESULT 14

US-09-878-178-717
 ; Sequence 717, Application US/09878178
 ; Patent No. US20020177552A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Secrist, Heather
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.527
 ; CURRENT APPLICATION NUMBER: US/09/878,178
 ; CURRENT FILING DATE: 2001-06-08
 ; NUMBER OF SEQ ID NOS: 2237
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 717
 ; LENGTH: 430
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-878-178-717

Alignment Scores:
 Pred. No.: 3,178-141 Length: 430
 Score: 143.00 Matches: 143
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.31% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x US-09-878-178-717 (1-430)

Qy 591 LeuArgGlnAsnThrAspGluHisAlaArgValMetSerTyrTyrAspTyrGlyTyrGln 610
 Db 2 CTAAGGCAAAATACAGATGAATGATGAGTAAATGCTTGTGGGATTAATGGCTATCAG 61
 Qy 611 IleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTyrAsnSerHis 630


```

RESULT 11
US-10-296-115-629
; Sequence 629, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 629
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-629

Alignment Scores:
Pred. No.: 3 51e-219 Length: 1114
Score: 217.00 Matches: 274
Percent Similarity: 98.21% Conservative: 0
Best Local Similarity: 98.21% Mismatches: 2
Query Match: 26.27% Indels: 5
DB: 16 Gaps: 0

US-10-028-384-2 (1-826) x US-10-296-115-629 (1-1114)
QY 1 MetAlaGluProSerAlaProGluSerHisHisSerSerSerLeuAsnSerSerProTyr 20
Db 211 ATGGCGAGCCCTCGGCCCGGAGAGCAAGCAAGTCGTCCTCAATCTGCCCGGG 270
QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyValadnGly 40
Db 271 AGTGCCCTCATGGCCCTGGAAACAGCCGCGCCAGCCAGCCGCGCCAGTGC 330
QY 41 AlaHisLysAlaAlaGlyGlyAlaProProLysProAlaPro---AlaGlyLeuSer 59
Db 331 GCGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 388
QY 60 GlyGlyLeuSerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrLeuPheLeu 79
Db 389 GGGGGGCTGTGCGAGCGCGCTGGTGGCAGTCGCTTCCTCTCCATCTCTCTCTG 448
QY 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValileArgPheGluSerleile 99
Db 449 GCCTGGCTTGGCGGCTTCAAGTCGCGGCTTTCGCGCTCATCCGCTTCGAAAGCATATC 508
QY 100 HisGluPheAspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPhe 119
Db 509 CACGAGTTCACCGCGGTGTTAACTATAGATCAACATCATCTTTCATCTCATCGGTTTC 568
QY 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTyrTrpProLeuGlyArgileValGly 139
Db 569 TATGAATTTTAAATTGGTTTATGAAGAGAGCATGATCCACTAGGAAGATAGTAGT 628
QY 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159
Db 629 GGTAAGTGTATACCGAGGTGATGATACCGCTGGCTTATTCATGGATTTTAAATACA 688
QY 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 689 TTGAACATACTGTTCACATAGAGAGCATATGTTGTTCTTTCACCACTTTTAGCGGC 748
QY 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199
Db 749 CTTACATCTATATCTATCTTCTGCTTACAGAGAACTTTGGAACCAAGAGCAGACTT 808
QY 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219

```

```

Db 809 TTAGCTGCTGTTTATTGCTATTGTACACGAGTACATATCTCGTCAAGTAGTGGATCC 868
QY 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrLeuTrpVallys 239
Db 869 TTTGATATGAAGCATTCGTTATTTTTCAGCTTCAGTTCACATATATTTATGGGTAATA 928
QY 240 SerValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMet 259
Db 929 TCTGTAAAAACCTGGGTCACTTTTGGCAATGTCGCTGCTTATCTATTTCTATATG 988
QY 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHis 276
Db 989 GTCTGCTGTTGGGTGGTTATGTTATTTATCATCATCTTATTTCCACTGCAT 1039

RESULT 12
US-10-264-049-630
; Sequence 630, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Barse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 630
; LENGTH: 2660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2634)..(2634)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2637)..(2637)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-630

Alignment Scores:
Pred. No.: 1 37e-196 Length: 2660
Score: 196.00 Matches: 296
Percent Similarity: 99.33% Conservative: 0
Best Local Similarity: 99.33% Mismatches: 1
Query Match: 23.73% Indels: 2
DB: 16 Gaps: 0

US-10-028-384-2 (1-826) x US-10-264-049-630 (1-2660)
QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetMetMet 549
Db 123 GGGTTAGCCCTAATATATAAAGCATGTGCACCATGTTGATGCTGATGCTATGATG 182
QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
Db 183 TTTGCTGCTCCACTGTACTGGGTCAAGCAATGCTCTACTAGTCCAAAGTAGTCCCTG 242
QY 570 AlaserTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
Db 243 GCCTCATACATCATGATGCGCAGCAATATCTTAGATGATTTTAGAGAGCTTACTTT 302
QY 590 TrpLeuArgGlnAenThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
Db 303 TGGCTAAGGCAATAATACAGATGAACATGCACGAGTATGTCCTGGTGGGATTATGGCTAT 362
QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnSer 629
Db 363 CAGATAGCTGAATGGCTAATAGAACTACGTTGGTGGATTAATACACCTGGAAATACAGC 422

```



```

Db 242 TGTGTGTTCTCTTGGCCCACTTTTAGGGCTTACATCTATATCTACTTCTCTGCTTACA 301
Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
Db 302 AGAAGAACTTTGGAACCAAGGACGAGGACTTTTAGCTGCTTTTATTTAGTCTATTGTACA 361
Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
Db 362 GGCTACATATCTCGTCTAGTAGTGGATCTTTGATATAGAGGCATCTCTATTGTGCA 421
Qy 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheIlePhe 249
Db 422 CTTGAGTTACATCTATTATGGTAAATCTGTAAACCTGGGTCAGTTTGTGGACA 481
Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
Db 482 ATGTGCTGCTCTTATCTTCTATATGTTCTCTGCTTGGGTTGTTATGATTATC 541
Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnThrGlyTyrSerIleArg 289
Db 542 ATCAATCTTATTCACATCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 601
Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
Db 602 GTCATACATAGCATATAGCACCTTTCTACATTTGGGTTTAAATATTATCAATGACATACCT 661
Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
Db 662 TTTGTGGGATTCGCCCAATCAGAACAGTGAACACATGGCAGCTGCGAGGTGCTTTTGA 721
Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnPhe 349
Db 722 TTGCTGCAAGCTTATGCTTTCTTGAGTATCTGAGAGACCGATTAAACAAACAGAGTTC 781
Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
Db 782 CAGACCCCTTTCTTTTGGGTGATCTACTAGCTGCGAGGTGCTGTGTCTTATGTCATC 841
Qy 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
Db 842 TATTGACTTATACAGCTTACATTTGCACCATGGAGTGGCAGGTTTATTCAATTTGGGAT 901
Qy 390 ThrGlyTyrAlaIleIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
Db 902 ACTGGGTATGCAAAATACACATTTCCAAATTAATGCAATCAGTGTCTGAGCATCAACCTACG 961
Qy 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
Db 962 ACTTGGGTGCTCTTCTTCTTGTATCTATATTTGTTGTTGTTGTTGTTGTTGTTGTTG 1021
Qy 430 TrpPheCysIleLeuAsnIleAsnAspGluArg 440
Db 1022 TGGTTCTGTCATCAAAATATCAACGATGAAGA 1054

```

RESULT 9

```

US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US20040185440A9
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020F1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE

```

```

; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 11 Gaps: 0

```

US-10-028-384-2 (1-826) x US-09-818-683-133 (1-1543)

```

Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
Db 2 TTCGCCGTCATCCGCTTCGAAGCATCATCCACGAGTTCGACCCGCTGGTTAACTATAGA 61
Qy 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
Db 62 TCAACACATCATCTTGCATCTCATGGTCTCTATGAATTTTAAATTTGGTTGATGAAGA 121
Qy 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
Db 122 GCATGTATTCCTACTAGGAGAAATAGTAGTGGTACTGTTTATCCCGAGGTGTATGATAACC 181
Qy 150 AlaGlyLeuIleHisIleTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
Db 182 GCTGGCCTTATTTCATTTGGATTTTAAATACATTTGAACATACTGTTTCACATAAGAGACGTA 241
Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
Db 242 TGTGTGTTCTTCTGTCACCAACTTTTAGCGGCTTACATCTATATCTATCTTCTGCTTACA 301
Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
Db 302 AGAGAACTTTGGAACCAAGGACGAGGACTTTTAGCTGCTGTTTATTGCTATTGTATGCA 361
Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
Db 362 GGCTACATATCTCGGTCAGTAGCTGATCCTTTGATATGAAGGCACTGCTATTATTGCA 421
Qy 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheIleThr 249
Db 422 CTTCACTTCACATCTATTATGGGTAAATCTGTAAACCTGGGTGCTGCTTTTGTGACA 481
Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
Db 482 ATGTGCTGCTCTTATCTCTATATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerIleArg 289
Db 542 ATCAATCTTATTCACATCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 601
Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
Db 602 GTCATACATAGCATATAGCACCTTTCTACATTTGGGTTTAAATATTATCAATGACATACCT 661
Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
Db 662 TTTGTGGGATTCGCCCAATCAGAACAGTGAACACATGGCAGCTGCGAGGTGCTTTTGA 721
Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnPhe 349
Db 722 TTGCTGCAAGCTTATGCTTTCTTGAGTATCTGAGAGACCGATTAAACAAACAGAGTTC 781
Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
Db 782 CAGACCCCTTTCTTTTGGGTGATCTACTAGCTGCGAGGTGCTGTGTCTTATGTCATC 841

```

```

; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-305-736-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-305-736-133 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleHleHisGluPheAspProTrpPheAsnTyrArg 109
DB 2 TTCGCCGTCATCCGCTTCGAAGGATCATCCAGGATTCGACCCGCTGGTTAACTATAGA 61
QY 110 SerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGATCATCTGCGGTTCTATGAATTTTAAATGGTTGATGAAGA 121
QY 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
DB 122 GCATGGTATCCACTAGGAAGAAATAGTAGGTGCTGTTTACCACGAGGTTGATGAACC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
DB 182 GCTGGCCCTTATTCATGATTTTAAATACATTCGACATACTGTTCACTAGACGATGA 241
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuThr 189
DB 242 TGTGTGTTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTACA 301
QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
DB 302 AGAGAACCTTTGGACCAAGAGAGAGAGACTTTAGCTGCTGTTTATTTGCTATGTACCA 361
QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GGCTACATATCTCGCTCAGTAGCTGGATCCTTTGATAATGAAGGCATTTGCTATTTGCA 421
QY 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheThrThr 249
DB 422 CTTGAGTTACATATACTATTATGGTAAATCTGTAAATACTGGTCAGTTTTCGACA 481
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
DB 482 ATGTGCTGCTGCTTATCTCTATTTCTATATGCTCTCTGCTGGGCTGTTATGATTTATC 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
DB 542 ATCAATCTTATTCACATGCAATGATTTGTTGTTACTGATGACAGATACACCAAGA 601
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATAGCACTTTCTACATTTGCGGTTTAAATATTCATTCAGATACCT 661
QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
DB 662 TTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGACCTGCGAGTGTCTTTGCA 721
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgIleThrLysGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTCTTCAGTATCTGAGACCGGATTAACAAACAAGAGTTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369

```

```

DB 782 CAGACCCCTTTCTTTTGGGTGTATCACTAGCTGAGGCTGCTGTCTCTTCTAGTGCATC 841
QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
DB 842 TATTTGACTTATACAGGTTACATTCGACCATGAGTGGCAGGTTTTTATTCATTGTGGAT 901
QY 390 ThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
DB 902 ACTGGGTATGCAAAATACACATTCCTCAATTTATTCATCATGCTCTGAGCATCAACTAGC 961
QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
DB 962 ACTTGGGTGCTTCTTCTTGTATACATATTTTGTATGATCTTCCACAGGCGCTT 1021
QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
DB 1022 TGGTTCCTGCATCAAAATATCAACGATGAAGA 1054

RESULT 8
US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US20030211472A1
; GENERAL INFORMATION:
; APPLICANT: Peng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-818-683-133 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleHleHisGluPheAspProTrpPheAsnTyrArg 109
DB 2 TTCGCCGTCATCCGCTTCGAAGGATCATCCAGGATTCGACCCGCTGGTTAACTATAGA 61
QY 110 SerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGATCATCTGCGGTTCTATGAATTTTAAATGGTTGATGAAGA 121
QY 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
DB 122 GCATGGTATCCACTAGGAAGAAATAGTAGGTGCTGTTTACCACGAGGTTGATGAACC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
DB 182 GCTGGCCCTTATTCATGATTTTAAATACATTCGACATACTGTTCACTAGACGATGA 241
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuThr 189

```

```

; PRIOR APPLICATION NUMBER: US 60/066,090
; PRIOR FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a.t.g. or c
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a.t.g. or c
US-09-974-879-133

Alignment Scores:
Pred. No.: 0 Length: 1543
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-974-879-133 (1-1543)
QY 90 PheAlaValIleArGpHeGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg 109
DB 2 TTGCGCGTCATCCGCTTCGAAAGCATCATCCAGGATCGACCGTGGTTAACTATAGA 61
QY 110 SerThrHisHisLeuAlaSerHisGlyPheThyGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGGCATCTCATCGGTTCTATCAATTTTAAATGGTTGATGAAAGA 121
QY 130 AlaTrpTrpProLeuGlyValGileValGlyThrValTrpProGlyLeuMetIleThr 149
DB 122 GCATGGTATCCACTAGGAGAGATAGTAGTGGTACTGTTTACCAGGGTTGATATACC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgaspVal 169
DB 182 GCTGCGCTTATTCATTCGATTTTAAATACATTAACATTAACATTAACATTAACATTAAC 241
QY 170 CysValPheLeuAlaProTrpPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
DB 242 TGTGTGTCCTTGGCAACCACTTTTAGCGGCTTACATCATCTATCTATCTTCTGCTTACA 301
QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
DB 302 AGAAGAACTTTGGAACCAAGGAGCAGGACTTTTACGCTGCTGTTTATTCCTATTGTACCA 361
QY 210 GlyTrpIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GGCTACATATCTCGTCAGTAGCTGGATCCTTTGATAATCAAGGCATCTCTATTTTGA 421
QY 230 LeuGlnPheThrThrTrpLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
DB 422 CTTGAGTTCACATACTATTATGGGTAAATCTGTAAACCTGGGTCAGCTTTTGGACA 481
QY 250 MetCysCysLeuSerTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPheIle 269
DB 482 ATGTGCTGCTCTTATCTATCTATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTrpSerLysArg 289
DB 542 ATCAATCTTATTCACCTGCATGATTTTGTGTTACTGATGATGATGATGATGATGATGATGAT 601
QY 290 ValTrpIleAlaTrpSerThrPheTrpIleValCylLeuLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATAGCACTTCTACATGTTGGGTTTAAATTAATTAATTAATTAATTAAT 661
QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329

```

```

DB 662 TTTGTGGGATCCAGCCATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTTGA 721
QY 330 LeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeuArgAspArgLeuThrLysGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTTGTGAGTATCTTCAGAGACCGGATTAACAAAAACAAGAGTTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
DB 782 CAGACCCCTTTCTTTTGGGTGTATCACTAGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCT 841
QY 370 TyrLeuThrTrpThrGlyTrpIleAlaProTrpSerGlyArgPheTrpSerLeuTrpAsp 389
DB 842 TATTTGACTTATACAGGTTACATTCACCATCGAGTGGCAGGTTTTTATTCATTGTTGGGAT 901
QY 390 ThrGlyTrpAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
DB 902 ACTGGGTATGCAAAATACACATTCATATATGATCAGTGTCTGAGCCTCAACCTACG 961
QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
DB 962 ACTTGGGTGCTTTCTTCTTGTATCTACATATCTTGTATGTACCTTCCACAGCAGGCTT 1021
QY 430 TrpPheCysLeuLysAsnIleAsnAspGluArg 440
DB 1022 TGGTTCTGCATCAAAATATCAACGATGAAGA 1054

RESULT 7
US-09-305-736-133
; Sequence 133, Application US/09305736
; Publication No. US20030088078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; EARLIER FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA31PI

; CURRENT APPLICATION NUMBER: US/10/264,237

; CURRENT FILING DATE: 2002-10-04

; PRIOR APPLICATION NUMBER: PCT/US01/16450

; PRIOR FILING DATE: 2001-05-18

; PRIOR APPLICATION NUMBER: US 60/205,515

; PRIOR FILING DATE: 2000-05-19

; NUMBER OF SEQ ID NOS: 2876

; SOFTWARE: PatentIn Ver. 3.1

; SEQ ID NO 412

; LENGTH: 1209

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1203)..(1203)

; OTHER INFORMATION: n equals a,t,g, or c

US-10-264-237-412

Alignment Scores:

Pred. No.: 0 Length: 1209
Score: 351.00 Matches: 351
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.49% Indels: 0
DB: 16 Gaps: 0

US-10-028-384-2 (1-826) x US-10-264-237-412 (1-1209)

QY 90 PheAlaValIleAArgPheGluSerIleHisGluPheAspProTrpPheAenTyrArg 109
DB 2 TTCGGCGTATCCGGTTCGAAGACATCATCCAGGATTCACCGCTGGTTAACTATAGA 61
QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrpPheAspGluArg 129
DB 62 TCAACACATCATCTTGCATCTCATGGTCTCTATGAATTTTAAATGGTTGATGAAGA 121
QY 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
DB 122 GCATGATATCCCTAGGAGAAATAGTAGTGGTACTGTTACCCAGGTTGTGTAAC 181
QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleAArgAspVal 169
DB 182 GCTGGCCTTATTCATTTGGATTTAAATACATTAACATTAACATTAACATTAACATTA 241
QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuThr 189
DB 242 TGTGTGTCCTTGCACCACTTTAGCGGCTTACATCTATCTACTTTCTCTGCTTACA 301
QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro 209
DB 302 AGAGAACCTTGAACCAAGGACGAGACTTTTAGCTGCTTTTATTGCTATTGTACCA 361
QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
DB 362 GGTACATATCTCGGTAGTACTGGATCCCTTTGATATGAGGCAATGCTATTATTGCA 421
QY 230 LeuGlnPheThrTyrTyrIleuTrpValYssSerValYssThrGlySerValPheTrpThr 249
DB 422 CTTCACTTCACATACTATTATGGTAAATCTGTAAATCTGGTCTGCTTTTGGACA 481
QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
DB 482 ATGTGCTGCTGTTATCCCTATTTCTATATGCTCTCTGTTGGGTGGTTATGATTATC 541
QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
DB 542 ATCAATCTTATTCCTCATGATTTTGTGTTTACTGATGACAGATACAGCAAAAGA 601
QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
DB 602 GTCTACATAGCATATAGCACTTCTACATTTGCGTTTATATATATCAATGCAATACCT 661

RESULT 6

US-09-974-879-133
; Sequence 133, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974, 879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239, 893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818, 683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305, 736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064, 911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064, 985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066, 094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/056, 100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066, 089
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066, 095
; PRIOR FILING DATE: 1997-11-17

QY 310 PheValGlyPheGlnProIleAArgThrSerGluHisMetAlaAlaGlyValPheAla 329
DB 662 TTTGTGGGATTCAGGCAATCAGAACAGTGAACACATGCAGCTCCAGGTCTCTTTGCA 721
QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuAspAspArgLeuThrIlyscGlnGluPhe 349
DB 722 TTGCTGCAAGCTTATGCTTTCTTTCAGATATCTGAGAGACCGATTAAACAAAACAGATTC 781
QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
DB 782 CAGACCCCTTTTCTTTTGGGTGATCACTAGCTGCAGGTCTGTCTTCTAGTGTCTATC 841
QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
DB 842 TATTGTACTTATACAGGTTACATTCACCATGGAGTGGCAGGTTTATTTCATTTGGGAT 901
QY 390 ThrGlyTyrAlaIlyHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
DB 902 ACTGGGTATGCAAAATATACACATTTCCAATTTATTCATCATCTGTCTGAGCATCAAGCTACG 961
QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
DB 962 ACTTGGGTGCTTCTTCTTTTGTATCTACATATCTTGTATGTATGATCTTCCACAGCGCTT 1021
QY 430 TrpPheCysIleIlyAsnIleAsnAspGluArg 440
DB 1022 TGGTTCTGCATCAAAATATCAACGATGAAGA 1054

Db 670 ATCAACAATTTCTCGATGGTTAGGTAGCTGAAGAGAAATCCCAAGACATTCGG 729
 Qy 692 GluSerAspTyrPheThrProGlnGlyCluPheArgValAspLysAlaGlySerProThr 711
 Db 730 GAAAGTGACTATTTTACCCCAACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACT 789
 Qy 712 LeuLeuAnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAsp 731
 Db 790 TTGTTGANTGCCATGATGATAAATGTCACTACTACAGATTTGGAGAAATGCAGCTGGAT 849
 Qy 732 PheArgThrProProGlyPheAspArgThrArgAsnAlaGluLeuLysAspLys 751
 Db 850 TTTCGTACACCCCAAGGTTTTCAGCCGACACGTAATGCTGAGATTCGAAATAGGACATT 909
 Qy 752 LysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgLysTyrLys 771
 Db 910 AAATTCAAACATTTGGAGAGAGCTTTACATCAGAACACTGGCTTGTAGGATATATAA 969
 Qy 772 ValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsnLysPhe 791
 Db 970 GTAAAGAGCCTGTATAACAGGAGACATTTAGATCACAACCTCGAGTCCACCAATTTTC 1029
 Qy 792 ProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrLysLysAsn 811
 Db 1030 CCAAAACAGAGAGTATTTGTCAAGAGACTACCAAAAGGAGGCTGGCTACATTAATAAT 1089
 Qy 812 LysLeuValPheLysLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 826
 Db 1090 AAGCTGGTGTAAAGAGGCAAGAAATATCTAAGAGAGACTGTT 1134

RESULT 4

US-10-106-698-330 Application US/10106698
 ; Sequence 330 Application US/10106698
 ; Publication No. US20030109690A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106.698
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 330
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1203)..(1203)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-106-698-330

Alignment Scores:

Pred. No.: 0 Length: 1209
 Score: 351.00 Matches: 351
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 42.49% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-2 (1-826) x US-10-106-698-330 (1-1209)

Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
 Db 2 TTTCGGCGTCAATCCGCTTTCGAAAGAGCATCATCCAGAGTTCGACCCGCTGGTTAACTATAGA 61
 Qy 110 SerThrHisHisLeuAlaSerHisGlyPheTyrCluPheLeuAsnTrpPheAspGluArg 129

Db 62 TCAACAATCTCTCGATCTCATGGTTCATGATGATTTTAAATTTGGTTTGTATGAAGA 121
 Qy 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149
 Db 122 GCATGGTATCCACTAGCAAGAAATAGTAGGTGTCTGTTTACCCAGGTTGATGATAC 181
 Qy 150 AlaGlyLeuIleHisTrpIleLeuAsnTrpLeuAsnIleThrValHisIleArgAspVal 169
 Db 182 GCTGGCTTATCTTGGATTTTAAATACATTTGAACATAACCTGTTTCAATAGAGACGTA 241
 Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 242 TGTGTGTCTCTGACCAAACTTTTAGCGGCTTATATATCTATATCTACTTCTCTGTTACA 301
 Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 Db 302 AGAAGACTTTGGAAACCAAGAGCAGGACTTTTAGCTGCTTGTATTATGCTATTGTACCA 361
 Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 362 GGCTACATATCTCGGTCACTAGCTGATCCTTTGATAATGAAGGCAATGCTATTTTTGA 421
 Qy 230 LeuGlnPheThrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
 Db 422 CTTGAGTTTCACTACTATTATTTAGGTAAATCTGTAAAACTGGGTCACTGTTTGTGACA 481
 Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
 Db 482 ATGTGCTGCTCTTATCTATATGCTCTCTGCTTGGGTGGTTATGTATTTATC 541
 Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
 Db 542 ATCAATCTTATCTCCACTGCTATGTTTGTGTTGTTACTGATGCAGAGATACAGAAAGA 601
 Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuLeuSerMetGlnIlePro 309
 Db 602 GTCTACATAGCATATAGCACTTTTCTACATGTTGGGTTTAAATATTATCAATGCAGATACCT 661
 Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 Db 662 TTTGTGGGATTCAGCCCATCAGAACAGTGAACATGCGCAGCTGCGAGGTGCTTTTGA 721
 Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrIleArgAspArgLeuThrLysGlnIlePhe 349
 Db 722 TTGCTGCAAGCTTATGCTTTCTTTCAGATATCTGAGAGACCCGATTAACAAAAACAAGATT 781
 Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 Db 782 CAGACCCCTTTTCTTTTGGGTGTATCACTAGCTGCGAGGTGCTGTTCTTAGTGTCTATC 841
 Qy 370 TyrIleThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
 Db 842 TATTTGACTTATACAGGTTTACATTCACCATCGAGTGGCAGGTTTTTATTTATTTGCGAT 901
 Qy 390 ThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 Db 902 ACTGGGTATGCAAAAATATACATTTCCATTTATGATGCTAGTGTCTGAGCATCAACCTACG 961
 Qy 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 Db 962 ACTTGGGTGTCTTTCTTCTTCTTGTATATATTTTGTATGTATGTATGTATGTATGTATGTAT 1021
 Qy 430 TrpPheCysLysLysAsnIleAsnAspGluArg 440
 Db 1022 TGGTCTGCTATCAAAAATATCAACGATGAAGA 1054

RESULT 5

US-10-264-237-412
 ; Sequence 412 Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.

Db 1369 AAATATCAACGATGAAAGAGATTTTCGCTCTGATGCGATCAGTGCTGTGACTTTC 1428
QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaI 474
Db 1429 CGAGCTGATGTTGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1488
QY 474 eAlaPheSerAsnValPheGluHisLeuValAspMetLeuValArgGluAsnProPr 494
Db 1489 CGCTTCTCCATGTTTGGAGCACTATTTGGGGGATGACATGAAAGGGAAACCCACC 1548
QY 494 oValGluAspSerSerAspGluAspAspLysArgAsnGln-GlyAsnLeuTyrAspLys 514
Db 1549 TGTGGAGGACACGACGTGATGAGGATGACAAAGAAAGAACCC-AGGAAACTTGTATGACAAG 1607
QY 514 IaGlyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr 533
Db 1608 CAGGTAAAGTGGAGAGCATGT-GACAGAGCAAGAGAA-ACCTGAAGAGGGCTTGGGCC 1665
QY 533 oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHI 553
Db 1666 CAACATCAAAAGCATTTGACCATGCTGATGCTCATGCTGATGATGTTTCGGTCCA 1725
QY 553 sCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAs 573
Db 1726 CTGACAGTGGGTACACAGCAAGCGCTACTCCAGTCCAGTGGTGGTCTGCTCTCTACAA 1785
QY 573 nHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGI 593
Db 1786 TCATGATGGTACCAGGAATATTATAGATGATTTTAGAGAACGCTACTTTTGGCTGAGACA 1845
QY 593 nAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGI 613
Db 1846 AAACAGGATGACACAGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1905
QY 613 yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLe 633
Db 1906 CATGCCCAACAGGACCACTCTGGTGGATAACAAACCTGGAAACACAGCCACATCGCACT 1965
QY 633 uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuA 653
Db 1966 GGTCCGAAAGATGCTCTCCATGAAACGGCCGCTATAAAATCATGAGGTC-CCTTG 2024
QY 653 sValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleA 673
Db 2025 ATGTCGATTTATGTTGTTATTTTCGAGGAGTGAATGGCTATTCGGGGACGATATCA 2084
QY 673 snLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgLus 693
Db 2085 ACAGTTCTCTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2144
QY 693 ex-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrIe 712
Db 2145 G-TGACTATTTTAC-CCAGCAGGAGAGTTCGAGTAGACAAAGCTGGTCTCTACTCT 2202
QY 712 uLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPh 732
Db 2203 GTTAAACTGCCTTATGATATAAAATGTCATACAGATTTGGAGAAATGAGCTAGATT 2262
QY 732 eaArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleY 752
Db 2263 TCGCACTCCCCAGGCTTTGACCAACAGCTAATGCTGAGATTGGAAATAAAGACATTAA 2322
QY 752 sPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysVa 772
Db 2323 ATTCAGCATTTGGAGGAGCTTTTACATCAGACCACTGCTGTCAGGATATATTAAGT 2382
QY 772 lLysAlaProAspAsnArgGluThrLeu 791
Db 2383 GAAGCACCTGACAAACAGGGAGACCTA 2410

RESULT 3

US-09-945-527-62

; Sequence 62, Application US/09945527
; Publication No. US20030055588A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20030055588A1el Nucleic Acid Molecules Encoding
; FILE OF INVENTION: Nucleic Acid and Protein Homologs
; FILE REFERENCE: 35800/237985
; CURRENT APPLICATION NUMBER: US/09/945,527
; CURRENT FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1767, 1771
; OTHER INFORMATION: n = A,T,C or G
US-09-945-527-62

Alignment Scores:

Pred. No.: 0 Length: 1828
Score: 355.00 Matches: 355
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.98% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x US-09-945-527-62 (1-1826)

QY 472 SerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAspAspMetLysArgGlu 491
Db 70 TCTGCATATTCCTTTTCAATGTTTGGACCATATTTGGGGATGACATGAAAGGGAA 129
QY 492 AsnProProValGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyr 511
Db 130 AATCCACCTGTGGAGGACAGCAGTGTAGGATGACAAAAGAAACCAAGAAAATTTGTAT 189
QY 512 AspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThrGluGluGlyLeu 531
Db 190 GATAGGCGAGGTAAAGTGAGGAAACATGCACTGACACAGGMAAAACTCGAAGGGATTA 249
QY 532 GlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAla 551
Db 250 GGCCTTAATATAAAAGCATTTGCACCATGTTGATGCTGATGCTGATGATGCTGCT 309
QY 552 ValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSer 571
Db 310 GTCCACTGTACCTGGGTCAACAGCAATGCTACTCTAGTCCAGTGTAGTCTCGGCTCA 369
QY 572 TyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeu 591
Db 370 TACAATCATGTGCGACAGGAATATCTTAGATGATTTTAGAGAAAGCTTTACTTTGGCTA 429
QY 592 ArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIle 611
Db 430 AGGCAAAATACAGATGACATGACGAGTAATGCTCTGTTGGGATTTAGCTATCAGATA 489
QY 612 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 631
Db 490 GCTGAAATGGCTAATAGAACTAGCTTGGTGGATAATAAAGCAACCTGGAAATTAACAGCACATA 549
QY 632 AlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr 651
Db 550 GCATCTGTGGGAAAGACTATGCTTTCTAATGAACAGACAGCCCTAATAATCATGAGGACT 609
QY 652 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
Db 610 CTAGATGTAGATTATGTTTGGTATTTTGGAGGGGTTATTGGCTATTCTGCTGATGAT 669
QY 672 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArg 691

```

Db      1981  TTTCGAGGGTATTGGCTATTCTGGTGGATCAACAAATTTCTCGGATGGTTAGG 2040
Qy      681  lIeAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Db      2041  ATAGCTGAAGAGAGACATCCCAAGACATTCGGGAAGTCACTATTATCCCCACAGGGA 2100
Qy      701  GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Db      2101  GAAATTCCTGTAGACAAAGCAGGATCCCTACTTTGTGAATTCCTTATGTATAAAATG 2160
Qy      721  SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
Db      2161  TCATACTACAGATTGGAGAAATCAGCTGGATTTCGTACACCCCGAGTTTGACCGA 2220
Qy      741  ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
Db      2221  ACACGTAATGCTGAGATTGGAATTAAGGACATTAATTCAAACATTTGGAAGAAGCITT 2280
Qy      761  ThrSerGluHisTyrLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
Db      2281  ACATCAGACACTGGCTGTGTAGATATATAAGTATAAAGCCTGNTAACAGGAGACA 2340
Qy      781  LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
Db      2341  TTAGATCAACAACCTCGAGTCACCAACATTTTCCCAAAACAGAAATATTGTCAAAAGA 2400
Qy      801  ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
Db      2401  ACTACCAAAAGAGAGCGTGGCTACATTAAATAAAGCTGTTTTTAAGAAGCAAGAAA 2460
Qy      821  IleSerLysLysThrVal 826
Db      2461  ATATCAAGAAGACTGTT 2478

```

RESULT 2

```

US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

```

Alignment Scores:

```

Pred. No.: 0 Length: 2710
Score: 365.00 Matches: 720
Percent Similarity: 98.09% Conservative: 0
Best Local Similarity: 98.09% Mismatches: 7
Query Match: 44.19% Indels: 14
DB: 15 Gaps: 0

```

US-10-028-384-2 (1-826) x US-10-028-384-3 (1-2710)

```

Qy      55  ProAlaGlyLeuSerGlyLeuSerGlnProAlaGlyTrpGlnSerLeuSerPhe 74
Db      230  CCCGCGGGGCTGTCCGGGGGCTTGTTCGACGCGCGGGTGGCAGTGTCTCTCTTC 289

```

```

Qy      75  ThrIleLeuPheLeuAlaTrpIleAlaGlyPheSerSerArgLeuPheAlaValIleArg 94
Db      290  ACCATCTCTCTCTGGCCCTGGCTGGCGGCTTCAGCTGGCGCTCTTCGCCGTATCCGC 349
Qy      95  PheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgSerThrHisLeu 114
Db      350  TTCGAGNCGATCATCCACGAGTTCGACCGCTGGTTTAACTATAGATCAACATCATCTT 409
Qy      115  AlaSerHisGlyPheTyrGluPheLeuAsnTyrPheAspGluArgAlaTyrProLeu 134
Db      410  GCATCTCATGGATCTATGAGTTTCTAAATTTGGTTTGAATGAAGAGCATGGTACCAC 469
Qy      135  GlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHis 154
Db      470  GGAAGATAGTGGGTGGCCCTTACCCAGGGTGTGATGAATGAACAGCTGGCTTATTCAT 529
Qy      155  TrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAla 174
Db      530  TGGATTTTAAATACATTTGAACATAACAGTTTCACATAAGAGATGTGTGTATTCCTGCA 589
Qy      175  ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsn 194
Db      590  CCAACTTTTAGCGGCTTACATCCATCTAGTTCCTGCTAACTAGAGAGACTGTGGAAAC 649
Qy      195  GlnGlyAlaGlyLeuLeuAlaLysPheIleAlaIleValProGlyTyrIleSerArg 214
Db      650  CAAGGAGCAGGACTTCTAGCTGCTTCAITTCATTCGTACCGAGGTACATATCTCGG 709
Qy      215  SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
Db      710  TCAGTGGCGGGAATCTTTGATTAATGAGGCATTTGCATTTTTCGCTTCAGTTCACTTAC 769
Qy      235  TyrIleuTrpValLysSerValLysThrGlySerValPheThrPheThrMetCysCysLeu 254
Db      770  TACTTATGGTAAAGTCTGTGAAGACCGGCTGTGTCTCGACAATGTGTGTGTGTGTG 829
Qy      255  SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIlePro 274
Db      830  TCATATTTCTACATGATGCTCTGGTGGGAGGTTATGTTCATCATCAACCTATCCCT 889
Qy      275  LeuHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
Db      890  CTCATGTTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 949
Qy      295  SerThrPhe-TyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGln 314
Db      950  AGCAC-TTTTGTACATTTGGGTTTATATATCCATGCGAGATACCTTTTGTGGGATTCA 1008
Qy      314  nProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTy 334
Db      1009  GCCAATCAGAAACAAGCGACATGCGAGCTGTCTTTGGGCTGTCTGCAAGCTTA 1068
Qy      334  rAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePh 354
Db      1069  CGCTTTTTCAGTATCTGAGAGACCGGTTGACAAACAGAGATTCAGACCCCTTTCTT 1128
Qy      354  eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrTh 374
Db      1129  TTTGGGTGTCTCACTAGCTGCAGGCGCTGTGTTCCTTGTGTGTGTGTGTGTGTGTGT 1188
Qy      374  rGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaTy 394
Db      1189  AGGTATATTTGCACCATGGAGTGGCAGGTTTATTTCACTATGGGATATCTGGGTATGCAA 1248
Qy      394  sIleHisIleProIleAlaSerValSerGluHisGlnProThrThrTrpValSerPh 414
Db      1249  AATACACATTCCAATTAATTCATCAGTGTCTGAACATCAGCCCTAGACATGCGGTCTCTT 1308
Qy      414  ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleTy 434
Db      1309  CTTCTTTGATCTACATATCTTGTATGTACCTTCCAGCAGGCGCTATGGTTCGATCA 1368
Qy      434  sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl 454

```

; OTHER INFORMATION:

US-10-028-384-1

Alignment Scores:

Pred. No.: 0
 Score: 826.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 15
 DB: 15

Length: 2481

Matches: 826

Conservative: 0

Mismatch: 0

Indels: 0

Gaps: 0

US-10-028-384-2 (1-826) x US-10-028-384-1 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisSerSerLeuAsnSerSerProTrp 20
 DB 1 ATGGCGAGAGCCCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCCCTCATGGCCCTGGGAAACAGCGCGCACGCCACCGCGCGCGCGCGCGCG 120
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 GGGCAAGAGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrLleLeuPheLeuAla 80
 DB 181 GGGCTGTGCGAGCG 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValLleArgPheGluSerLleHis 100
 DB 241 TGGCTGTGCG 300
 QY 101 GluPheAspProTrpPheAsnTrpArgSerThrHisLysLeuAlaSerHisGlyPheTrp 120
 DB 301 GAGTTTCGAGCCCGTGTAACTATAGATCAACACATCATCTTCGATCGTCTCTAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProGluArgLleValGlyGly 140
 DB 361 GAATTTTAAATGGTTTGTATGAAGAGCATGGTATCCATCAGGAAAGATAGTAGGT 420
 QY 141 ThrValTrpProGlyLeuMetIleThrAlaGlyLeuLleHisTrpLleLeuAsnThrLeu 160
 DB 421 ACTGTTTACCAGGGGTTCATGATAACCGCTGGCTTATTCATTGGAATTTTAAATAC 480
 QY 161 AsnLleThrValHisLysArgAspValCysValPheLeuAlaProTrpPheSerGlyLeu 180
 DB 481 AACATAACTGTTTCACATAGAGACGATGTGTGTCTTCCTTGACCAACTTTTAGCGG 540
 QY 181 ThrSerLleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 DB 541 ACATCTATATCTACTTTCTCTTACAGAGAACTTTTGAACCAAGGAGGAGGAGGAG 600
 QY 201 AlaAlaCysPheLleAlaAlaValProGlyTrpLleSerArgSerValAlaGlySerPhe 220
 DB 601 GCTGTGTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTGAGTGTGATGCTCT 660
 QY 221 AspAsnGluGlyLleAlaLlePheAlaLeuGlnPheThrTrpTrpValLysSer 240
 DB 661 GATAATGAAGGCATTTGCTATTTTTCAGCTTCAGTTCATATCTTTATGGTAAATCT 720
 QY 241 ValLysThrGlySerValPheTrpThrMetCysCysLeuSerTrpPheTrpMetVal 260
 DB 721 GTAAAAACTGGGTGAGTTTGTATTTATCATCAATCTTATTCCTGCTATCTATATG 780
 QY 261 SerAlaTrpGlyGlyTrpValPheLleAlaAsnLeuLleProLeuHisValPheValLeu 280
 DB 781 TCTGCTGGGTGGTGTATTTATTTATCATCAATCTTATTCCTGCTATCTATATG 840
 QY 281 LeuLeuMetGlnArgTrpSerLysArgValTrpLleAlaTrpSerThrPheTrpLleVal 300
 DB 841 TTACTGATGCAGAGATACAGCAAGAGTCTTACATAGATATAGCACTTTCTACATG 900

QY 301 GlyLeuLleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTTTAATATATCAATGCAGATACCTTTTGTGGATTCCAGCAATCAGAAAGTGAA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeu 340
 DB 961 CACATGGCAGTGGGTGCTTTGTCATGCTCAAGCTATGCTTCTTCAGTATCTG 1020
 QY 341 ArgAspArgLeuThrTrpGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGACCGATTAACAAAAAAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValLleTrpLeuThrTrpThrGlyTrpLleAlaProTrp 380
 DB 1081 CGAGTGTGTTTCTTCTTAGTCACTATTTGACCTTATACAGGTTCATCTGCACATG 1140
 QY 381 SerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysLleHisLleProLle 400
 DB 1141 AGTGGCAGGTTTATTCATTGCTGGATCTGGGTATGCAAAAAATACACATTCATTA 1200
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisLle 420
 DB 1201 GCATCAGTGTCTGAGCATCAACCTACAGTGGGTGCTTTCTCTTTCATCTACATAT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysLleLysAsnLleAsnAspGluArg 440
 DB 1261 CTTGTATGTACTTCCCGCAGCGCTTGGTTCGTCATCAAAAAATACACGATGAAAGA 1320
 QY 441 ValPheValAlaLeuTrpAlaLleSerAlaValTrpPheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTTGTTGCTCTATATGCAATCAGTGTCTCTCTTTGCTGGAGTGTATGGTGG 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaLleAlaPheSerAsnValPhe 480
 DB 1381 ATGTTGACTTTCAGTCCAGTCGTGTATGCTGCTGCAATTCCTTTTCAATGTTTT 1440
 QY 481 GluHisTrpLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCAGTATTTGGGGGATGACATGAAAAGGAAAAATCCACCTGTGGAGACAGCATGT 1500
 QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTrpAspLysAlaGlyLysValArgLysHis 520
 DB 1501 GAGGATGACAAAAAAGAACCAAGGAAATTTGATGATAAGCGAGGTAAAGTGAGAAAC 1560
 QY 521 AlaThrGluGlnGlyThrGluGluGlyLeuGlyProAsnLleLysSerLleValThr 540
 DB 1561 GCAACTGACAGGAAAAAACTGAAGAGGATTAAGGCCCTTAATAAAAAAGCATTTGCAC 1620
 QY 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 DB 1621 ATGTTGATGCTGATGCTTATGATGATGTTTGTCTGCTCCACTGTACCTGGGTCA 1680
 QY 561 AlaTrpSerSerProSerValValLeuAlaSerTrpAsnHisAspGlyThrArgAsnLle 580
 DB 1681 GCCTACTCTAGTCCAGTGTAGTCTCGGCTCTATCAATCATGATGTCACAGGATATC 1740
 QY 581 LeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 DB 1741 TTAGATGATTTAGAGAGCTTACTTTGGCTAAGGCAAAATACAGATGAAATGACATG 1800
 QY 601 ValMetSerTrpTrpAspTrpGlyTrpGlnLleAlaGlyMetAlaAsnArgThrLeu 620
 DB 1801 GTATGCTTGGTGGGATTAAGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860
 QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisLleAlaLeuValGlyLysAlaMetSerSer 640
 DB 1861 GTGGATAATAACCTGGAATTAACAGCCACATAGCACTGGTGGAAAAAGCTATGCTCT 1920
 QY 641 AsnGlnThrAlaAlaTrpLysLleMetArgThrLeuAspValAspTrpValLeuValLle 660
 DB 1921 AATGAACAGCAGGCTATAAAATCATGAGGACTTAGATGATGATGATGTTTGGTTAT 1980
 QY 661 PheGlyGlyValLleGlyTrpSerGlyAspAspLysPheLeuTrpMetValArg 680

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 ; Search time 1039.95 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPESKHSLNSPWP.....GYIKNLVFKGKKKSKTKTV 826

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame-p2n model -DEV=xlh
-Q/cgn2_1/USPTC spool/US10028384/runat_14122004_131648_6892/app query.fasta_1.3740
-DB=PublishedApplications NA -OFMT=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1
-LOOPCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-NAXLEN=200000000 -USER=US10028384@cgn 1 1 2124 @runat_14122004_131648_6892
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-Fgapop=6 -Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 826 | 100.0 | 2481 | 15 | US-10-028-384-1 |
| 2 | 365 | 44.2 | 2710 | 15 | US-10-028-384-3 |
| 3 | 355 | 43.0 | 1828 | 10 | US-09-945-527-62 |
| 4 | 351 | 42.5 | 1209 | 15 | US-10-106-698-330 |
| 5 | 351 | 42.5 | 1209 | 16 | US-10-264-237-412 |
| 6 | 351 | 42.5 | 1543 | 10 | US-09-974-879-133 |
| 7 | 351 | 42.5 | 1543 | 10 | US-09-305-736-133 |
| 8 | 351 | 42.5 | 1543 | 10 | US-09-818-683-133 |
| 9 | 351 | 42.5 | 1543 | 11 | US-09-818-683-133 |
| 10 | 351 | 42.5 | 1543 | 16 | US-10-621-401-133 |
| 11 | 217 | 26.3 | 1114 | 16 | US-10-296-115-629 |
| 12 | 196 | 23.7 | 2660 | 16 | US-10-264-049-630 |
| 13 | 166 | 20.1 | 500 | 9 | US-09-998-598-1643 |
| 14 | 143 | 17.3 | 430 | 9 | US-09-878-178-717 |
| 15 | 143 | 17.3 | 430 | 13 | US-10-046-935-717 |
| 16 | 143 | 17.3 | 430 | 14 | US-10-146-502-717 |
| 17 | 133 | 16.1 | 558 | 14 | US-10-052-283-433 |
| 18 | 119 | 14.4 | 387 | 16 | US-10-276-774-173 |
| 19 | 117 | 14.2 | 616 | 9 | US-09-879-536-332 |
| 20 | 108 | 13.1 | 485 | 10 | US-09-918-995-11283 |
| 21 | 104 | 12.6 | 743 | 10 | US-09-945-527-63 |
| 22 | 68 | 8.2 | 637 | 15 | US-10-002-631C-183 |
| 23 | 68 | 8.2 | 640 | 15 | US-10-002-631C-179 |
| 24 | 54 | 6.5 | 474 | 14 | US-10-052-283-454 |
| 25 | 51 | 6.2 | 393 | 16 | US-10-276-774-323 |
| 26 | 44 | 5.3 | 312 | 9 | US-09-796-692-8870 |
| 27 | 44 | 5.3 | 312 | 14 | US-10-040-862-8870 |
| 28 | 44 | 5.3 | 312 | 16 | US-10-057-475B-8870 |
| 29 | 44 | 5.3 | 312 | 16 | US-10-154-884B-8870 |
| 30 | 44 | 5.3 | 312 | 17 | US-10-764-324-8870 |
| 31 | 42 | 5.1 | 2417 | 15 | US-10-028-384-7 |
| 32 | 40 | 4.8 | 352 | 9 | US-09-878-178-1032 |
| 33 | 40 | 4.8 | 352 | 13 | US-10-046-935-1032 |
| 34 | 40 | 4.8 | 352 | 14 | US-10-146-502-1032 |
| 35 | 34 | 4.1 | 483 | 10 | US-09-918-995-23896 |
| 36 | 33 | 4.0 | 1848 | 15 | US-10-128-714-2139 |
| 37 | 33 | 4.0 | 1969 | 15 | US-10-128-714-1139 |
| 38 | 33 | 4.0 | 2232 | 15 | US-10-128-714-7139 |
| 39 | 33 | 4.0 | 2603 | 15 | US-10-128-714-6139 |
| 40 | 33 | 4.0 | 3969 | 15 | US-10-128-714-139 |
| 41 | 33 | 4.0 | 4603 | 15 | US-10-128-714-5139 |
| 42 | 21 | 2.5 | 154 | 10 | US-09-991-936-450 |
| 43 | 20 | 2.4 | 502 | 17 | US-10-767-701-6513 |
| 44 | 19 | 2.3 | 2256 | 15 | US-10-032-585-6323 |
| 45 | 17 | 2.1 | 887 | 16 | US-10-424-599-77697 |

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPARTIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)

APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 3090
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-3090

Alignment Scores:
Pred. No.: 4,428-07 Length: 594
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.94% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-248-796A-3090 (1-594)

Qy 92 VallileArgpHeGluSerllelelleHiscLupheAspProtrpPheasn 107
Db 151 GFGATTGCAATTGAAAGTATTATTTCATGAATTCGATCCTTGGTTCAAT 198

Search completed: December 14, 2004, 23:43:35
Job time : 170.937 secs

GENERAL INFORMATION

APPLICANT: Keith Weinstock et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

RESULT 15
US-09-248-786A-3090
; sequence 3090, Application US/09248796
; Patent No. 6747137
; GENERAL INFORMATION:

```

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:
Pred. No.: 1.53e-30 Length: 1660
Score: 39.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.72% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-158 (1-1660)
QY 652 LeuAspValAspTyrValIleuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
Db 673 CTGACGTGGACTAGCTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCGGCGATGAT 614
QY 672 IleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
Db 613 ATCAACAAGTTCTCTGGATGTCGGAATTCGGAGGAGAGCATCCCAAGGACATT 557

RESULT 8
US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 1.53e-30 Length: 1660
Score: 39.00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.72% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-15440 (1-1660)
QY 652 LeuAspValAspTyrValIleuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 671
Db 673 CTGACGTGGACTAGCTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCGGCGATGAT 614
QY 672 IleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 690
Db 613 ATCAACAAGTTCTCTGGATGTCGGAATTCGGAGGAGAGCATCCCAAGGACATT 557

RESULT 9
US-09-270-767-12331
; Sequence 12331, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12331
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12331

Alignment Scores:
Pred. No.: 1.9e-21 Length: 900
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.63% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-12331 (1-900)
QY 80 AlaTrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIle 99
Db 160 GCCTGGCTGGCGGATTTCTCTCGCTCTCGCGTCATCGTTTCGATCGATTATC 219
QY 100 HisGluPheAspProTyrPheAsnTyrArg 109
Db 220 CATGAGTTTGATCGGTGTTCAACTACCGG 249

RESULT 10
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Alignment Scores:
Pred. No.: 1.57e-18 Length: 560
Score: 27.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.27% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-1318 (1-560)
QY 445 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 464
Db 423 CTGTACGCCATCAGTCGGTTTACTTCGCTGGTGTGATGCGTTTGATGTGACCCCTC 364
QY 465 ThrProValValCysMetLeu 471
Db 363 ACCCGCGTGGTGTGCATGCTG 343

RESULT 11
US-09-270-767-1660/c
; Sequence 1660, Application US/09270767

```

QY 286 TyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSer 305
Db 63 TACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAATATATCA 122
QY 306 MetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 325
Db 123 ATGCAGATACCTTTTGTGGGATTCAGCAATCAGAACATGGAACATGGCAGCTGCA 182
QY 326 GlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrIleuArgAspArgLeuThr 345
Db 183 GGTGCTTTTGCAATGCTCTCAAGCTATGCTTTCTTTCAGTATCTGAGACCCGATTACA 242
QY 346 LysGln 347
Db 243 AAACAA 248

RESULT 4

US-09-513-999C-21090
; Sequence 21090, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Malne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21090
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21090

Alignment Scores:
Pred. No.: 7,586-72 Length: 245
Score: 78.00 Matches: 78
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.44% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-513-999C-21090 (1-245)

QY 185 ThrPheLeuThrArgGluLeuTyrAsnGlnGlyAlaGlyLeuLeuAlaCysPhe 204
Db 3 ACTTCTGCTTACAGAGACTTTGAAACCAAGAGCAGGACTTTTAGCTGCTGTTT 62
QY 205 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspGluGly 224
Db 63 ATTGCTATTGACAGGCTACATATCTCGTTCAGTAGCTGATCCCTTGTATGAGGC 122
QY 225 IleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLysSerValLysThrGly 244
Db 123 ATTGCTATTTTGTACITTCAGTTACATATCTTTATGGTAAATCTGTAAACCTGGG 182
QY 245 SerValPheThrThrMetCysCysLeuSerTyrPheTyrMetValSerAla 262
Db 183 TCAGTTTTTGGCAATGTGCTGCTTATCTCTTATTTCTATATGCTCTCTGCT 236

RESULT 5

US-09-270-767-158
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger, et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:
Pred. No.: 1,176-33 Length: 1660
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.08% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-158 (1-1660)

QY 599 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 618
Db 854 GCTCGCTTATGCTTTGGTGGGATTACGGATACCGATAGCGGSAATGCAACAGAACG 913
QY 619 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 638
Db 914 ACCTAGTGGATATATATACGTGGACATATAGTCATAGCGCTGTTGGCAAGCAATG 973
QY 639 SerSer 640
Db 974 TCTTCA 979

RESULT 6

US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No.: 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 1,176-33 Length: 1660
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.08% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-270-767-15440 (1-1660)

QY 599 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 618
Db 854 GCTCGCTTATGCTTTGGTGGGATTACGGATACCGATAGCGGSAATGCAACAGAACG 913
QY 619 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 638
Db 914 ACCTAGTGGATATATATACGTGGACATATAGTCATAGCGCTGTTGGCAAGCAATG 973
QY 639 SerSer 640
Db 974 TCTTCA 979

RESULT 7

QY 562 TyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeu 581
Db 2 TACTCTAGTCCCAAGGTAGTCTCGCTCATACATCATGATGSCACACAGGAATATCTTA 61
QY 582 AspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgVal 601
Db 62 GATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGSCACGAGTA 121
QY 602 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuVal 621
Db 122 ATGTCCTGGTGGAATATAGCTATCAGATAGCTGGATGGCTTAATAGAACTACGTTGGTG 181
QY 622 AspaenAnthrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsn 641
Db 182 GATAATAACACCTGGAATTAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTCTAAT 241
QY 642 GluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValIleValIlePhe 661
Db 242 GAAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATTTT 301
QY 662 GlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArgIle 681
Db 302 GGAGGGGTATTTGGCTATTTCTGGTGATGATATCAACAAATTTCTCTGGATGGTTAGGATA 361
QY 682 AlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGlu 701
Db 362 GCTGAAGGAGACATCCCAAGACATTCGGGAAAGTGCATATTTTACCCACAGGGAGAA 421
QY 702 PheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSer 721
Db 422 TTCGCTGTAGACAAAGCAGGATCCCTACTTTGTTGAATTGCCTTAATGATATAAATGCTCA 481
QY 722 TyrTrpArgPheGlyGluMetGln 729
Db 482 TACTACAGATTGGAGAAATGCAG 505

RESULT 2

US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333

GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(616)
; OTHER INFORMATION: n = A, T, C or G

US-09-328-111-332

Alignment Scores:

Pred. No.: 5,73e-112 Length: 616
Score: 117.00 Matches: 117
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.16% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-2 (1-826) x US-09-328-111-332 (1-616)

QY 208 ValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIle 227
Db 5 GTACCAGGCTACATATCTCGTCAGTAGCTGATCTTTGATATGAAGGCAATTCCTATT 64
QY 228 PheAlaLeuGlnPheThrTyrTyrLeuTrpValIysSerValIysThrGlySerValPhe 247
Db 65 TTTCGACTTCAGTTTACATACATATTTATGGGAAAAATCTGTAAAAAATCTGGTTCAGTTT 124
QY 248 TrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrVal 267
Db 125 TGGACAAATGTGCTGCTTATCCCTATTTCTATATATGCTCTGCTTGGGTGGTTATGTA 184
QY 268 PheIleIleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSer 287
Db 185 TTATCATCAATCTTATTTCCACTGCATGATTTTGTGTTTACTGATCCAGAGATACAGC 244
QY 288 LysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGln 307
Db 245 AAAAGAGTCTACATAGCATATAGCACTTTTACATTTGCGGTTTAAATATTATCATGCGAG 304
QY 308 IleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 324
Db 305 ATACCTTTTGTGGGATTCAGCAATCAGAAACAAGTGAACACATGCGCAGCT 355

RESULT 3

US-09-513-999C-1438
; Sequence 1438, Application US/09513999C
; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1438
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 54-248
; US-09-513-999C-1438

Alignment Scores:

Pred. No.: 5,41e-76 Length: 250
Score: 82.00 Matches: 82
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.93% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-513-999C-1438 (1-250)

QY 266 TyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArg 285
Db 3 TAATGTTTATCATCAATCTTATTCACATGATGTTTGTGTTACTGATGCGAGAGA 62

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 160.937 Seconds
(without alignments)
3648.073 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPSKHKSLSNPPW.....GYKNKLVPFKKISKKTIV 826

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=xlh
-Q/cgn2_1/USPTC_spool/US10028384/runat_14122004_131646_6916/app_query.fasta.1.3740
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=Oligo.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human0.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @cgn 1 1 259 @runat_14122004_131646_6916 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -HEADS=1 -XGAPEXT=60 -XGAPEXT=60 -XGAPEXT=60
-XGAPEXT=7 -XGAPEXT=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|---------------------|
| 1 | 168 | 20.3 | 507 | US-09-513-999C-1965 | Sequence 1965, App |
| 2 | 117 | 14.2 | 616 | US-09-328-111-332 | Sequence 332, App |
| 3 | 82 | 9.9 | 250 | US-09-513-999C-1438 | Sequence 1438, App |
| 4 | 78 | 9.4 | 245 | US-09-513-999C-21090 | Sequence 21090, A |
| 5 | 42 | 5.1 | 1660 | US-09-270-767-158 | Sequence 158, App |
| 6 | 42 | 5.1 | 1660 | US-09-270-767-15440 | Sequence 15440, A |
| C 7 | 39 | 4.7 | 1660 | US-09-270-767-15440 | Sequence 15440, App |
| C 8 | 39 | 4.7 | 1660 | US-09-270-767-15440 | Sequence 15440, A |
| C 9 | 30 | 3.6 | 900 | US-09-270-767-12331 | Sequence 12331, A |
| C 10 | 27 | 3.3 | 560 | US-09-270-767-1318 | Sequence 1318, App |
| C 11 | 27 | 3.3 | 560 | US-09-270-767-16600 | Sequence 16600, A |
| 12 | 20 | 2.4 | 133 | US-09-270-767-28082 | Sequence 28082, A |

| | | | | | | |
|----|----|-----|------|---|---------------------|-------------------|
| 13 | 19 | 2.3 | 867 | 4 | US-09-248-796A-3089 | Sequence 3089, Ap |
| 14 | 17 | 2.1 | 2157 | 4 | US-09-614-221A-318 | Sequence 318, App |
| 15 | 16 | 1.9 | 594 | 4 | US-09-248-796A-3090 | Sequence 3090, Ap |
| 16 | 15 | 1.8 | 487 | 3 | US-09-385-982-213 | Sequence 213, App |
| 17 | 14 | 1.7 | 302 | 4 | US-09-313-294A-6869 | Sequence 6869, Ap |
| 18 | 13 | 1.6 | 914 | 4 | US-09-270-767-12856 | Sequence 12856, A |
| 19 | 13 | 1.6 | 1386 | 4 | US-09-270-767-11648 | Sequence 11648, A |
| 20 | 12 | 1.5 | 268 | 4 | US-09-313-294A-4834 | Sequence 4834, Ap |
| 21 | 12 | 1.5 | 291 | 4 | US-09-313-294A-4834 | Sequence 4834, Ap |
| 22 | 12 | 1.5 | 299 | 4 | US-09-313-294A-4834 | Sequence 4834, Ap |
| 23 | 11 | 1.3 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| 24 | 11 | 1.3 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| 25 | 11 | 1.3 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| 26 | 11 | 1.3 | 307 | 4 | US-09-736-457-375 | Sequence 375, App |
| 27 | 11 | 1.3 | 307 | 4 | US-09-736-457-375 | Sequence 375, App |
| 28 | 11 | 1.3 | 307 | 4 | US-09-736-457-375 | Sequence 375, App |
| 29 | 11 | 1.3 | 307 | 4 | US-09-614-124B-375 | Sequence 375, App |
| 30 | 11 | 1.3 | 307 | 4 | US-09-614-124B-375 | Sequence 375, App |
| 31 | 11 | 1.3 | 307 | 4 | US-09-614-124B-375 | Sequence 375, App |
| 32 | 11 | 1.3 | 307 | 4 | US-09-614-124B-375 | Sequence 375, App |
| 33 | 11 | 1.3 | 307 | 4 | US-09-671-325-588 | Sequence 588, App |
| 34 | 11 | 1.3 | 307 | 4 | US-09-671-325-588 | Sequence 588, App |
| 35 | 11 | 1.3 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 36 | 11 | 1.3 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 37 | 11 | 1.3 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 38 | 11 | 1.3 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 39 | 11 | 1.3 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 40 | 11 | 1.3 | 487 | 4 | US-09-702-705-1655 | Sequence 1655, Ap |
| 41 | 11 | 1.3 | 487 | 4 | US-09-702-705-1655 | Sequence 1655, Ap |
| 42 | 11 | 1.3 | 487 | 4 | US-09-614-124B-1655 | Sequence 1655, Ap |
| 43 | 11 | 1.3 | 487 | 4 | US-09-614-124B-1655 | Sequence 1655, Ap |
| 44 | 11 | 1.3 | 487 | 4 | US-09-671-325-1655 | Sequence 1655, Ap |
| 45 | 10 | 1.2 | 109 | 4 | US-09-658-824-1655 | Sequence 1655, Ap |
| | | | | | US-09-270-767-28703 | Sequence 28703, A |

ALIGNMENTS

RESULT 1

US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513, 999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Alignment Scores:
Pred. No.: 5,08e-165 Length: 507
Score: 168.00 Matches: 168
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.34% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x US-09-513-999C-1965 (1-507)

| | | | |
|----|-----|--|-----|
| QY | 761 | ThrSerGluHisTrpLeuValArgIleTyrLysValIysAlaProAspAsnArgGluThr | 780 |
| Db | 663 | ACAICAGACACTGGCTTGTAGCATATATAAGTAAAGCACCTCATACAGGGAGACA | 722 |
| QY | 781 | LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys | 800 |
| Db | 723 | TTAGATCACAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTAATTTGTCAAAGAAG | 782 |
| QY | 801 | ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys | 820 |
| Db | 783 | ACTACCAAAAGAGCGTGGCTACATTAAATATAGCTGCTTTTAAAGAGGCAAGAA | 842 |
| QY | 821 | IleSerLysLysThrVal | 826 |
| Db | 843 | ATATCTAAGAAGACTGTT | 860 |

Search completed: December 14, 2004, 22:36:14
Job time : 9206.03 secs

```

QY 558 ThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThr 577
Db 119 ACACGAATGCTACTCTAGTCCCAAGTGTAGTCTGCGCTCAACATCATGATGACACC 178
QY 578 ArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGlu 597
Db 179 AGGAATATCTTAGATGATTTTAGAGAGCTTACCTTTGGCTAAGGCAAAATACAGATGAA 238
QY 598 HisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 617
Db 239 CATGACAGATGATGCTCTGGTGGATATGCTGATCATGATGATGATGATGATGATGATGAT 298
QY 618 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 637
Db 299 ACTACGTTGGTGGATAATAACACCTGGATAAACAAGCACACATAGCACTGGTGGGAAAGCT 358
QY 638 MetSerSerAsnGluThrAlaAlaTyrIleValMetArgThrLeuAspValAspTyrVal 657
Db 359 ATGCTCTCTATGAACACGACGCTATTAATCATGAGGACTCTAGATGATGATGATGATGAT 418
QY 658 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 677
Db 419 TTGGTTATTTTGGAGGGTATTTGGCTATTTCTGCTGATGATATCAACAAATTTCTCTGG 478
QY 678 MetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThr 697
Db 479 ATGGTTAGATAGCTGAGAGGAGACATCCCAAGACATTCGGAAAGTGAATTTTACC 538
QY 698 ProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMet 717
Db 539 CCACAGGAGAGATTCGCTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTCCTTATG 598
QY 718 TyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGly 737
Db 599 TATAAATGTCTACTACAGATTTGGAGAAATGCGAGCTGGATTTTGTACACCCCGAGT 658
QY 738 PheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGlu 757
Db 659 TTTGACCAAGACACGCTAATGCTGAGATTCGAAATGGAATTAAGACATTAATTAACATTTG 718
QY 758 GluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsn 777
Db 719 GAAGCTTTATCATCAGAACACTGCTGTTAGGATATATTAAGTATAAAGCACCCTGATAAC 778
QY 778 ArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeu 797
Db 779 AGGAGACATTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAAATTTTG 838
QY 798 SerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLys 817
Db 839 TCNAGAGAGACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGGTTTAAAGAA 898
QY 818 GlyLysLysIleSerLysLysThrVal 826
Db 899 GCGAAGAAATATCTAAGAGAGCTGTT 925

RESULT 15
AX882932
LOCUS 2284 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 17837 from Patent EP1074617.
ACCESSION AX882932
VERSION AX882932.1 GI:40037833
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesizing full-length cDNA and their use
Patent: EP 1074617-A 17837 07-FEB-2001;

```

Research Association for Biotechnology (J2)

FEATURES

Location/Qualifiers

1. 2284
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

CDS

3. 7863
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAE91783.1"
/db_xref="GI:40037834"
/translation="MLMLMLLMVFAVHCTWTSNAYSSPVSVLASYNHDGTRNLLDDF
REAYFLRQNTDEHAFVMSWDYQYIAGMANRTLLVDNTWNSHIALVAGKAMSNE
TAAYKIMRLDVIDYLVIFGGVIGSGDDINKFLMWRIAEGEHPKDRSDYFTPOG
EPFVDKAGSPTLLNCLMRYKVSRYRFGEMCLDPTPPGDFRTRNAEIGNKD:KFHLE
AFTSEHMLVRIYKVRAPDNRELDHKPRVTNIFPKQKYLSSKTTKRGYIKNKLVEK
KGKISKIV"

ORIGIN

Alignment Scores:
Pred. NO.: 7,72e-297 Length: 2284
Score: 286.00 Matches: 286
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 34.62% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x AX882932 (1-2284)

```

QY 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
Db 3 ATGTTGATGCTGATGCTATTGATGATGTTTGTCTCCACTGTACCTGGGTCAACAGCAAT 62
QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
Db 63 GCCTACTCTAGTCCAGTAGTGTGCTGGCTCATCAATCATGATGAGGACCAAGGAATATC 122
QY 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
Db 123 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATGCACGA 182
QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 620
Db 183 GTATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 242
QY 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
Db 243 GTGATAATAACACCTCGAATAACAGCCACATAGCACCTGGTGGGAAAGCTATGCTTCT 302
QY 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
Db 303 AATGAACACAGCAGCCATATAAATCATGAGGACTCTAGATGTAGATTATGTTTGGTTATT 362
QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
Db 363 TTTGGAGGGTATTATGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGATGTTAGG 422
QY 681 IleAlaGluGlyLysIleProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Db 423 ATAGCTGAAGGAGAAATCCCAAGACATTCGGGAAAGTGAATATTTTACCCCAAGGGA 482
QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Db 483 GAATTCGCTGATAGAACAAAGCAGGATCCCTACTTTGTTGAAATTTGCTTATGATAAATG 542
QY 721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
Db 543 TCATATCTACAGATTTGGAGAAATGACAGCTGGGATTTTCGTACACCCCAAGGTTTGGCCGA 602
QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
Db 603 ACACGTAATGCTGAGATTGGAAATAAGGACATTAATAATCAACAAATTTTGAAGAGACCTTT 662

```


Db 721 GACATTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 780
 Qy 770 TTTTCTTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 789
 Db 781 TTTTCTTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 840
 Qy 790 TTTTCTTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 809
 Db 841 TTTTCTTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 900
 Qy 810 TTTTCTTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 826
 Db 901 TTTTCTTAAATTCAGAGAGCCCTTATACAGAACACCTGGCTTGTAGGATA 951

RESULT 12

AK075380

LOCUS

DEFINITION

Homo sapiens cDNA PSEC0070 fis, clone NT2RP2001508, moderately

similar to OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT.

ACCESSION

AK075380

VERSION

AK075380.1

KEYWORDS

oligo capping, fis (full insert sequence).

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1

Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayaashi, K.,

Ishii, S., Saito, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,

Nakamura, Y., Nagahara, K., Sugano, S. and Isogai, T.

HRI human cDNA sequencing project

Unpublished

2 (bases 1 to 2510)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,

Genomics Laboratory, 1332-3 Yana, Kisarazu, Chiba 252-0812, Japan

(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass

sequencing, clone selection and full insert sequencing; Helix

Research Institute (supported by Japan Key Technology Center etc.);

cDNA library construction; Institute of Medical Science, University

of Tokyo, Laboratory of Genome Structure, Human Genome Center.

Location/Qualifiers

1. 2510

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="NT2RP2001508"

/cell_lines="NT2"

/clone_lib="NT2RP2"

/note="Cloning vector: pME18SPL3-mRNA from NT2 neuronal

precursor cells after 2-weeks retinoic acid (RA)

induction"

94. .954

/note="unnamed protein product"

/codon_start=1

/protein_id="BAC11581.1"

/db_xref="GI:22761429"

/translation="MLMLMLMFAVHCTVNTSNVSSPSVVLASVNHGTRNILDFF

REAVFWLRQNDHARVMSWDYQIQAGNKKITLVLDNNTWNNHIALVGVKMSNE

TAATKRLDLDVILVIFGVIGVIGDDINKFLMVRVIAEGHPKIDRESDFTPQG

EPRVDKAGSPFLNCLMYKMSYVYFQMDFFTPFQFTRNAEIGNDKIRPKLEE

AFTGEHMLVRIYKVKAFDNRETLDHPKRVNTNIPFKQKYLKSKTKTKRKYIKNKLVFK

KGKKISKTV"

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Length: 2510
 Matches: 317
 Conservative: 0
 Mismatches: 0

Query Match: 38.38%

DB: 9

Indels: 0

Gaps: 0

US-10-028-384-2 (1-826) x AK075380 (1-2510)

Qy 510 LeuTyrAspLysAlaGlyLysVala:GlyHisAlaThrGluGlnGluLysThrGluGlu 529

Db 1 TTGTATGATTAAGGCAGGTAAAGTGAAGAACATGCACTGAACAGGAAAAAACTGAAGAG 60

Qy 530 GlyLeuGlyProGlnLysSerLLeValThrMetLeuMetLeuMetLeuMetMet 549

Db 61 GGATAGGCCCTTAATAATAAAGACATGTGCACCATGTTGATGCTGATGCTATTGATGATG 120

Qy 550 PheAlaValHisCysThrTrrValThrSerAsnAlaTyrSerSerProSerValValLeu 569

Db 121 TTTCTCTCCACTGTACTGGGTCAACAAGCAATGCCTACTCTAGTCCAAAGTGTAGTCTGT 180

Qy 570 AlaSerTyrAsnHisAspGlyThrArgAsnLeuAspAspPheArgGluAlaTyrPhe 589

Db 181 GCCTATCAATCAATGATGACCCAGGAATATCTTAGATGATTTTAGAGAGCTTACTTT 240

Qy 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrrTrpAspTyrGlyTyr 609

Db 241 TGGCTAAGGCAAAATACAGATGAACATGCACGAGTAACTCTTGTGGTGGATTTATGGCTAT 300

Qy 610 GlnIleAlaGlyMetAlaSerArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 629

Db 301 CAGATAGCTGGATGGCTAATAGAACTCTGTTGGTGGATTAATAACACCTGGGAATACAGC 360

Qy 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649

Db 361 CACATAGCACTGGTGGGAAAAGCTATGTTCTTAATGAAACAGCAGCCTATAAATCATG 420

Qy 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669

Db 421 AGGACTCTAGATGTAGATTAAGTTTGGTATTTTGGAGGGGTATTGGCTATTCTGTGT 480

Qy 670 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 689

Db 481 GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAAGAGAGAACATCCCCAAGAC 540

Qy 690 IleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySer 709

Db 541 ATTGGGAAAGTGACTATTTTACCCACAGGGGAAATTCGGTGTAGACAAACAGAGATCC 600

Qy 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTrrTyrArgPheGlyGluMetGln 729

Db 601 CCTACTTTGTTGAATTTGGCTTATGTATAAAATGTTCATACAGATTTGGAGAAATGCAG 660

Qy 730 LeuAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyValAsnLys 749

Db 661 CTGGATTTTCTGTACACCCCGAGTTTGGACCGNACACGTAATGCTGAGATTGGAAATAAG 720

Qy 750 AspIleLysPheLysHisLeuGluAlaPheThrSerGluHisTrpLeuValArgIle 769

Db 721 GACATTAATTCACAAATTTTGAAGAAGCGCTTTACATCAGAACACCTGGCTTTTAGGATA 780

Qy 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsn 789

Db 781 TATTAAGTAAAGCACTGATTAACAGGAGACATTAGATCAAAACCTCGAGTCCACCAAC 840

Qy 790 IlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIle 809

Db 841 ATTTTCCAAAACAGAGATTTTGTCAAAAGAGACTACCAAAAGGAGCGGTGGCTACATT 900

Qy 810 LysAsnLysLeuValPheLysLysGlyLysValIleSerLysLysThrVal 826

Db 901 AAAAATPAGCTGCTTTTAAAGAAAGCAAGAAATATCTTAAGAGACTGTT 951

RESULT 13

AX017997

LOCUS

DEFINITION

Sequence 232 from Patent WO9946375.

2503 bp DNA linear PAT 07-SEP-2000

QY 530 TrrLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
 Db 241 TGGCTAAGCGCAAAATACAGATGAACATGCAGGAGTAATGCTTGGTGGGATTATGGCTAT 300
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsnSer 529
 Db 301 CAGATAGCTGGAATGGCTTAATAGAACATACCTTGGTGAATAAATACCACTGGAAATACAGC 360
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
 Db 361 CACATAGCACTGGTGGGAAGCTATGCTTCTTAATGAACACAGCAGCTTAAATCATG 420
 QY 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyClyValIleGlyTyrSerGly 569
 Db 421 AGGACTCTAGATGATAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGT 480
 QY 670 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 689
 Db 481 GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAGGAGAAACATCCCAAGAC 540
 QY 690 IleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySer 709
 Db 541 ATTCGGGAAGTGATATTTTACCCACAGGAGAAATTCGCTAGACAAAGCAGATCC 600
 QY 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
 Db 601 CCTACTTTGTTGAAATTTCTGATATATATAAATGTCTACTACAGATTTGAGAAATGTCAG 660
 QY 730 LeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
 Db 661 CTGGATTTTCTGATACACCCCAAGGTTTGGACGACAGTAATCTGAGATTTGGAATAAG 720
 QY 750 AspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIle 769
 Db 721 GACATTAATTCMAACATTTGGAAGAGCCTTTACATCAGAACACTGGCTTGTAGGATA 780
 QY 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAsnHisLysProArgValThrAsn 789
 Db 781 TATAAAGTAAGAAGCACCTTGATTAACAGGAGACATTAGATCACAAGCTCGAGTCACCAAC 840
 QY 790 IlePheProLysGlnLysTyrLeuSerLysLysThrLysArgLysArgGlyTyrIle 809
 Db 841 ATTTTCCCAACACAGAGATTTTGTCAAGAAGACTACCAAAAGGAAGGTGGCTACATT 900
 QY 810 LysAsnLysLeuValPheLysGlyLysLysLysLysSerLysLysThrVal 826
 Db 901 AAAATAAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAGAAGACTGTT 951

RESULT 11
 BD123520 2510 bp DNA linear PAT 18-SEP-2002
 LOCUS
 DEFINITION Secretory protein or membrane protein.
 ACCESSION BD123520
 VERSION BD123520.1 GI:23218465
 KEYWORDS JP 2002017376-A/29.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
 Hayashi,K.
 SECRETORY PROTEIN OR MEMBRANE PROTEIN
 Patent: JP 2002017376-A 29 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017376-A/29
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253173
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU
 SUGIYAMA,
 KOJI HAYASHI
 PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10, C12P21/02, C12Q1/68//C12P21/08, C12N15/00, C12N5/00 CC
 Secretory protein or membrane protein
 FH Key Location/Qualifiers
 FT CDS (94)..(951).
 FEATURES
 source 1..2510
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores: 0 Length: 2510
 Pred. No.: 317.00 Matches: 317
 Score: 317.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.38% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x BD123520 (1-2510)
 QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThrGluGlu 529
 Db 1 TTGTATGATAAGGCGAGTAAAGTGAAGAAACATGCAACTGAACAGGAAAAAAGTCAAGAG 60
 QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMet 549
 Db 61 GGATTAGCCCTAATAATAAAGCAATGTCACCATGTTGATGCTGATGCTATTGATGATG 120
 QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
 Db 121 TTGTGCTGCCACTGTAACCTGGGTGCAAGCAATGCTACTCTAGTCCAAGGTAGTCTG 180
 QY 570 AlaserTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyr-Phe 589
 Db 181 GCTCATCATCATCATGATGGCACCAGGAATATCTTAGATGATTTAGAGAGCTTACTTT 240
 QY 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
 Db 241 TGGCTAAGGCAAAATACAGATGAACATGCAGGAGTAATGCTTGGTGGGATTATGGCTAT 300
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 629
 Db 301 CAGATAGCTGGAATGGCTTAATAGAACATGCTTGGTGGATATAACACCTGGAAATACAGC 360
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
 Db 361 CACATAGCACTGGTGGGAAGCTATGCTTCTTAATGAACACAGCAGCTTAAATCATG 420
 QY 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
 Db 421 AGGACTCTAGATGATAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGT 480
 QY 670 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 689
 Db 481 GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAGGAGAAACATCCCAAGAC 540
 QY 690 IleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySer 709
 Db 541 ATTCGGGAAGTGATATTTTACCCACAGGAGAAATTCGCTAGACAAAGCAGATCC 600
 QY 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
 Db 601 CCTACTTTGTTGAAATTTCTGATATATATAAATGTCTACTACAGATTTGGAATAAG 660
 QY 730 LeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
 Db 661 CTGGATTTTCTGATACACCCCAAGGTTTGGACGACAGTAATCTGAGATTTGGAATAAG 720
 QY 750 AspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIle 769

```

FEATURES             Location/Qualifiers
     source           1..2546
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.:          0          Length:      2546
Score:             323.00      Matches:      323
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:       39.10%      Indels:      0
DB:                6          Gaps:         0

US-10-028-384-2 (1-826) x BD063986 (1-2546)
QY 504 LysArgGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523
Db 1 AAAGAAACCAAGGAAATTTGTATGATAGGACGATTAAGTGAGGAAACATGCAACTGAA 60
QY 524 GlnGluThrGlnGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543
Db 61 CAGGAAAAAAGCTGAAGAGGATAGGCCCTAATAATAAAAGCATTTGTCACCATGTTGATG 120
QY 544 LeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563
Db 121 CTGATGCTATTGATGATGTTTCTGTCTCCACTGTCACCTGGGTCAAGCAATGCTACTCT 180
QY 564 SerProSerValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAsp 583
Db 181 AGTCCAGGTAGTCTCGCCCTCATCATCATGATGCCACCAAGGAATATCTTAGATGAT 240
QY 584 PheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspLysHisAlaArgValMetSer 503
Db 241 TTTAGAGAGCTTACTTTTGCTAAGGCAAAATACAGATGAACATGCACGAGTAATGCT 300
QY 604 TrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
Db 301 TGGTGGGATATTGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAAT 360
QY 624 AsnThrTrpAsnAsnSerHisAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
Db 361 AACACCTGGATTAACAGCCCATAGCACTGCTGGGAAAGCATATGCTTCTTAATGAACA 420
QY 644 AlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGly 663
Db 421 GCAGCCTATAAAATCATGAGGACTCTAGATGATGATTTATGTTTGGTTATTTTGGAGGG 480
QY 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlu 683
Db 481 GTTATGGCTATTCTGGGATGATATCAACAATTTCTCGATGGTTAGGATAGCTGAA 540
QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
Db 541 GGAGAAATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGAGAAATTCCT 600
QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
Db 601 GTAGCAAAAGCAGGATCCCTTCTTGTGAATTCCTTATGATTAATAATGTCATACTAC 660
QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
Db 661 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCACCGAACACAGTAAT 720
QY 744 AlaGluIleGlyAsnLysAspIleLysPheIleHisLeuGluAlaPheThrSerGlu 763
Db 721 GCTGAGATTGAATTAAGGACATTAATTAACATTTTGAAGAAGCCCTTACATCAGAA 780
QY 764 HisTrpLeuAlaGlyTyrLysValIleAlaProAspAsnArgGluThrLeuAspHis 783
Db 781 CACTGGCTTGTAGGATATAAAGTAAAGACCTGATAAACAGGAGACATTAGATAC 840

```

```

QY 784 LysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrLys 803
Db 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATATTGTCAAGAGACTACCAA 900
QY 804 ArgLysArgGlyTyrIleLysAsnLysLeuValPheLysGlyLysLysLysSerLys 823
Db 901 AGGAAGCGTGGCTACATTAAATAATAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAG 960
QY 824 LysThrVal 826
Db 961 AAGACTGTT 969

RESULT 10
AX136135
LOCUS AX136135 2510 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 57 from Patent EP1067182.
ACCESSION AX136135
VERSION AX136135.1 GI:14272543
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
 1 Ota, T., Isozaki, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and
Hayashi, K.
 2 Secretory protein or membrane protein
 3 Patent: EP 1067182-A 57 10-JAN-2001;
 4 Helix Research Institute (JP)
 5 Location/Qualifiers
 6 1..2510
 7 /organism="Homo sapiens"
 8 /mol_type="unassigned DNA"
 9 /db_xref="taxon:9606"
10 94..954
11 /note="unnamed protein product"
12 /codon_start=1
13 /protein_id="CAC39710.1"
14 /db_xref="GI:14272544"
15 /translation="MLMLMLMFAVHCTWTTSNAYSSPSVVLASYNHGTGRLNLLDDF
16 REAFVRLONTDEHARVMSWDYGYIAGMANRTTLVDNNTNNSHIALYVSKAMSSNE
17 TAYKKMLRLDVLVIFGGVIGYSGDDINKFLMWRVIAEGHPKDIRESDYTFPGQ
18 EFRVDKAGSFTLLNLMYSYRFGEMQLDFRTPGPDRTNRNABIGNKDIKPKHLEE
19 AFTSEHMLVRIYKVKAPDNRELDHKPRVTNIFPKQYLSKKTTRKRGYIKNKLVEK
20 KKKISKSTV"
CDS
 1..2510
 2 /note="unnamed protein product"
 3 /codon_start=1
 4 /protein_id="CAC39710.1"
 5 /db_xref="GI:14272544"
 6 /translation="MLMLMLMFAVHCTWTTSNAYSSPSVVLASYNHGTGRLNLLDDF
 7 REAFVRLONTDEHARVMSWDYGYIAGMANRTTLVDNNTNNSHIALYVSKAMSSNE
 8 TAYKKMLRLDVLVIFGGVIGYSGDDINKFLMWRVIAEGHPKDIRESDYTFPGQ
 9 EFRVDKAGSFTLLNLMYSYRFGEMQLDFRTPGPDRTNRNABIGNKDIKPKHLEE
10 AFTSEHMLVRIYKVKAPDNRELDHKPRVTNIFPKQYLSKKTTRKRGYIKNKLVEK
11 KKKISKSTV"
ORIGIN
Alignment Scores:
Pred. No.:          0          Length:      2510
Score:             317.00      Matches:      317
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:       38.38%      Indels:      0
DB:                6          Gaps:         0

US-10-028-384-2 (1-826) x AX136135 (1-2510)
QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnLysThrGluGlu 529
Db 1 TTGTATGATAGGACGATTAAGTGAGAAACATGCACTGAACAGGAAAAAACTGAAGAG 60
QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMet 549
Db 61 GGATTAGGCCCTAATAATAAAAGCATTTGTCACCATGTTGATGCTGATCTATTGATGATG 120
QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValVal 569
Db 121 TTTGCTGTCTCACCTGTCACCTGGGTCAAGCAATATGCTACTCTAGTCCAAGTGTAGTCTG 180
QY 570 AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
Db 181 GCCTCATACATCATCATGATGCACCAAGGAATACTTAGATGATTTTATAGAGAGCTTACTT 240

```


QY 712 uLeuAsnCysLeuMetTyrIysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPh 732
 Db 2203 GTTAACTGCCCTTATGATATAAAATGTCATACAGATTGGAGAAATCGAGTAGATT 2262
 QY 732 eArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleY 752
 Db 2263 TCGCACTCCCCAGCTTTGACCGAACACGTAATGCTGAGATTGGAAATTAAGACATTA 2322
 QY 752 sPheLysHisLeuGluLulAlaPheThrSerGluHisTrpLeuValArgIleTyrLysVa 772
 Db 2323 ATTCAAGCATTTGGAGGAAGCTTTTACATCAGAGCACTGGCTTGTGAGGATATAAAGT 2382
 QY 772 llyeAlaProAspAsnArgGluThrLeu 781
 Db 2383 GAAAGCACCTGACACACAGGGAGACACTA 2410
 RESULT 8
 AX099510
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS
 Jacobs, K., Mccoy, J.M., Lavallie, E.R., Collins-Racie, L.A., Evans, C.,
 Merberg, D., Treacy, M., Bowman, M.R., Spaulding, V. and Agostino, M.J.
 TITLE
 Secreted proteins and polynucleotides encoding them
 JOURNAL
 Patent: WO 01/9988-A 150 22-MAR-2001;
 Genetics Institute, Inc. (US)
 FEATURES
 source
 1..2546
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2546
 Score: 323.00 Matches: 323
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 29.10% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x AX099510 (1-2546)
 QY 504 lYsArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523
 Db 1 AAAAGAACCAAGGAAATTTGTATGATAAGCGAGTTAAAGTGAGGAAACATGCAACTGAA 60
 QY 524 GlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543
 Db 61 CAGGAAAAAATGAGAGGGGATTAGGCCCTTAATAAAAGCAATGTCACCATGTGTGAG 120
 QY 544 LeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563
 Db 121 CTGATGCTATTGATGATGTTTGTGCTCCACTGTACCTGGGTCAACAGCAATGCTACTCT 180
 QY 564 SerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAsp 583
 Db 181 AGTCCAAGTGTAGTCTGGCTCTATCAATCATGATGCGACCAAGGATATCTTAGATGAT 240
 QY 584 PheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603
 Db 241 TTTAGAGAGCTTACTTTTGCTTAAGGCAAAATACAGATCAACATGCAAGTAAATGTCT 300
 QY 604 TrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623

Db 301 TGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAAGAACTACGTTGGTGAATAA 360
 QY 624 AsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
 Db 361 AACACCTGGAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTCTTCTTAATGAACA 420
 QY 644 AlalaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGly 663
 Db 421 CGAGCCTTAATAATCATGAGGACTCTAGATGATGATTAATGTTTGGTTATTTTGGAGG 480
 QY 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlu 683
 Db 481 GTTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGATGGTGTAGGATAGCTGAA 540
 QY 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 Db 541 GGAGAACATCCCAAGACATTCGGGAAGATGACTATTTTACCCACAGGGAGAATTCGT 600
 QY 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 Db 601 GTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGATATAAAATGCTACTACT 660
 QY 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
 Db 661 AGATTTGGAGAATGCGCTGGATTTTGTACACCCCGAGTTTGTACCGAACACGTAAT 720
 QY 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPheThrSerGlu 763
 Db 721 GCTGAGATTGGAATTAAGGACATTAATAATTCAAACATTTTGGAGAAGCCCTTTACATCA 780
 QY 764 HisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThrLeuAspHis 783
 Db 781 CATGCTCTTGTAGGATATTAAGTAAGCACTGATTAACGGGAGACATTAAGATCAC 840
 QY 784 LysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrLys 803
 Db 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATATTTGTCAAGAGAGACTACCAA 900
 QY 804 ArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLysIleSerLys 823
 Db 901 AGGAAGCGTGGCTACATTAATAATAAGCTGCTTTTAAAGAAAGCAAGAAATATCTAAG 960
 QY 824 LysThrVal 826
 Db 961 AAGACTGTT 969
 RESULT 9
 BD063986
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 BD063986
 Secreted proteins and polynucleotides encoding them.
 BD063986
 BD063986.1 GI:22609589
 JP 2001506848-A/10.
 unidentified
 unclassified
 unclassified.
 1 (bases 1 to 2546)
 REFERENCE
 AUTHORS
 Treacy, M., Mccoy, J.M., Lavallie, E.R., Racie, L.A., Merberg, D.,
 Treacy, M., Spaulding, V. and Agostino, M.J.
 TITLE
 Secreted proteins and polynucleotides encoding them
 JOURNAL
 Patent: JP 2001506848-A 10 29-MAY-2001;
 GENETICS INSTITUTE INC
 COMMENT
 PN JP 2001506848-A/10
 PD 29-MAY-2001
 PF 12-DEC-1997 JP 1998525996
 PR 13-DEC-1996 US 08/766263
 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
 DAVID MERBERG,
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
 C12N15/12, C12N5/10, C07K14/47, A61K38/17
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Location/Qualifiers.

/organism="Mus musculus"
/no_type="mRNA"
/db_xref="taxon:10090"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2710
Score: 365.00 Matches: 720
Percent Similarity: 98.09% Conservative: 0
Best Local Similarity: 98.09% Mismatches: 7
Query Match: 44.19% Indels: 14
DB: 6 Gaps: 0

US-10-028-384-2 (1-826) x AX799084 (1-2710)

QY 55 ProAlaGlyLeuSerGlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPhe 74
DB 230 CCGCGGGGGCTGTCGGGGGGCTTGTGCGAGCGCGGGTGGAGTGTGCTCTCTTC 289
QY 75 ThrileLeuPheLeuAlaTrpLeuAlaGlyPheSerArgLeuPheAlaValIleArg 94
DB 290 ACCATCCCTCTTCCTGGCGCTGCTGGCGGCTTCAGCTCGCGGCTCTTCGCGGTCATCCGC 349
QY 95 PheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgSerThrHisIleSteu 114
DB 350 TTCGAGAGCATCATCCAGAGCTTCGACCGGTTGATTAAGTATAGATCAACATCATCTT 409
QY 115 AlaSerHisGlyPheTrpGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeu 134
DB 410 GCATCTCATGATTCATGATGTTCTAAATTTGGTTTGTATGAAGAGCATGGTACCCACTG 469
QY 135 GlyArgIleValGlyGlyThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHis 154
DB 470 GGAAGAAATAGTGGGTGGCCCGTTTACCCAGGTTGATGATACAGCTGGCCTTATTCAT 529
QY 155 TrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysPheLeuAla 174
DB 530 TGGATTTTAAATACATTAACATACAGTTACATAGAGATGTTGTGTTATTCCTTGCA 589
QY 175 ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsn 194
DB 590 CCAACTTTTAGCGGCTTACATCCATCATCTACGTTCTCTTAAGTACAGAGTGTGGAAC 649
QY 195 GlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValProGlyTrpIleSerArg 214
DB 650 CAAGAGAGAGACTTCTAGCTGCTTGTCTTCAATGCTATCTGACAGGATACATATCTCG 709
QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrp 234
DB 710 TCAGTGGCGGATCCTTTGATAATGAAGGCAATGGCAATTTTGGGCTTCAGTTCATCTAC 769
QY 235 TyrLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCysCysLeu 254
DB 770 TACTTATCGGTAAGTCTGTGAGACCGGGTCTGTCTGGAACAATGTGCTGCTCTG 829
QY 255 SerTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPheIleIleLeuLeuPro 274
DB 830 TCATATTTCTACATGGTCTCTGCGTGGGAGGTTATGTTTCATCATCAACCTCATCCCT 889
QY 275 LeuHisValPheValLeuLeuLeuMetGlnArgTrpSerLysArgValTrpIleAlaTrp 294
DB 890 CTCATGTTGTGTCTGCTGATGACAGAGTACAGCAAGAGTCTACATAGCATAT 949
QY 295 SerThrPhe-TyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheG1 314
DB 950 AGCAC-TTGTGTACATTTGGGTTTAAATATATCCATGCATACCTTTTGGGATTTCA 1008
QY 314 nProfilegthrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTrp 334
DB 1009 GCCAATCAGACACAGCAGACATGGCAGCTGCAGGTTCTTTGGCGCTGTCGACGCTTA 1068
QY 334 rAlaPheLeuGlnTrpLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePh 354

DB 1069 CGCTTTTTCAGTATCTGAGACCGGTTGACAAAACAGGAGTTCCAGACCTTTCTT 1128
QY 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTrpLeuThrTrpTh 374
DB 1129 TTTGGGTGTCTCAGTCTGAGCGGCTGTGTCTTCTGTAGTGTCTATCTGACATACAC 1188
QY 374 xGlyTrpIleAlaProTrpSerGlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLy 394
DB 1189 AGGTATATTGACCATGGAGTGGCAGGTTTATTTCATCTATGGGATCTGGGTATGCAAA 1248
QY 394 sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpTrpValSerPh 414
DB 1249 AATACACATTCCAATTAATTCATCAGTCTGAACATCAGCCTACGATGCTGTCTT 1308
QY 414 ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLy 434
DB 1309 CTTCCTTGAATCATATCTTGTATGTACCTTCCAGCAGGCTTATGGTCTGATCAAA 1368
QY 434 sAsnIleAsnAspGluArgValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAl 454
DB 1369 AAATATCAACGATGAAGAAGATATTGTGCTCTGTATGCGATCAGTGTGTGTCTTTC 1428
QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIle 474
DB 1429 CGGAGTGTGTGGCTGCTGCTGATCTGACCTGACCCCGGTCGTCGATGCTGTGGCCAT 1488
QY 474 eAlaPheSerAsnValPheGluHisTrpLeuGlyAspAspMetLysArgGluAsnProPr 494
DB 1489 CGCTTCTCCAAATGTTTTTGACACTATTGCGGATGACATGAAGGAGGAAACCCACC 1548
QY 494 oValGluAspSerSerAspGluAspAspLysArgAsnGln-GlyAsnLeuTrpAspLysA 514
DB 1549 TGTGGAGGACAGCAGTGATGAGGATGACAAAAGAAACCC-AGSAAACTTGTATGACAAGG 1607
QY 514 laglyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr 533
DB 1608 CAGTAAAGTGAGGAAGCATGT-GACAGAGCAAGAGAA-ACCTGAAGAGGGGCTTGGGGCC 1665
QY 533 oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValIle 553
DB 1666 CAACATCAAAAGCAATTTGACCATGCTGATGCTCATGCTCTGATGATGTTTGGCGTCCA 1725
QY 553 sCysThrTrpValThrSerAsnAlaTrpSerSerProSerValValLeuAlaSerTrpAs 573
DB 1726 CTGACGCTGGGTCAACAGCAAGCCCTACTCCAGTCCAAGTGGTGGTCTGTGCTCTCAAA 1785
QY 573 nHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTrpPheTrpLeuArgG1 593
DB 1786 TCATGATGTTACCGAGGAATATATAGATGATTTTAGAGAAGGCTACTTTTGGCTGAGACA 1845
QY 593 nAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaG1 613
DB 1846 AAACACGGATGAACACCGCCGGGTCTGTCGTGGTGGGACTACGGCTATCAGATTGCTGG 1905
QY 613 yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLe 633
DB 1906 CATGGCCCAACAGCAGCCACTCTGGTGGTAACCAACACCTGGAACAACAGCCATCGCACT 1965
QY 633 uValGlyLysAlaValSerSerAsnGlnThrAlaAlaTrpIleMetArgThr-LeuA 653
DB 1966 GGTGCGAAAAGCTATGTTCTCAATGAACCGCCGCTATATAAATCATGAGTCCCTTG 2024
QY 653 spValAspTrpValLeuValIlePheGlyGlyValIleGlyTrpSerGlyAspAspIleA 673
DB 2025 ATGTTCGATTATGTTGGTTATTTTCGGAGGAGTGTGATTTGCTTATTCGGGGACGATA 2084
QY 673 snLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluS 693
DB 2085 ACAAGTTCTCTGGATGGTCAGGATAGCTGAAGGGGAGGATCCCAAGACATCCGGGAG 2144
QY 693 er-AspTrpPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe 712
DB 2145 G-TGACTATTTCAC-CCAGCAGGGAGAGTTCCGAGTAGACAAAGCTGGGTCTCTCTACTCT 2202

FEATURES

Location/Qualifiers
1. .1664
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEMBA1006430"
/tissue="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/note="cloning vector: pME189FL3"
28. .1362
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11076.1"
/db_xref="GI:22760123"
/translation="MAPSPAPESKHSKLSNPSWGLMALGNHRHGHGPGACQAHKA
AGAAPKPPAPAGLGGSLQOPAGWQSLFTILFLAWLAGFSRLFAVIRFESIIHEF
DPWFNYSTRPLASHGFEFLNWFDERAYPLGRIVGGTVFPLMITAGLIHWILNTL
NITVIRDCVFLAFTPTSGSLTSISTELLRELWNOGALLAACFIAIVPGVIRSVAG
SFDNEGAIYALQFTYLVKVSFKTGVFWTCCCLSVFVMSAWGVFVNIPLH
VFVLLMRYLTKQEFOTFLGYSLAAGAVLSVLYLVTVGYIAWKGKRFVSLWDGY
AKIHPIIASVSEHQFTWVSPPFDLHILVCFPAGLMFCIKNIDRVFGA"

CDS

568 ACATCTATATCTACTTCTCTCTTCAAGAGAACTTTGGAAACCAAGAGAGAGACTTTTA 627
201 AlaAlaCysPheAlaAlaValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
628 GCTGCTGTTTATTGCTATTGCTACAGGCTACATATCTCGGTGAGTAGCTGATCTTT 687
221 AspAsnGluGlyLeuAlaIlePheAlaLeuGlnPheThrTyrLeuTyrValIleSer 240
688 GATAATGAAGGAGGATGCTATTATTTGCACTTCACTATCATATTTATGAGGTAATCT 747
241 ValIleThrGlySerValPheThrMetCysCysLeuSerTyrPheTyrMetVal 260
748 GTAAAGCTGGGTGAGTCTTTTGGCAATGCTGCTCTTATCTATCTATATGATGTC 807
261 SerAlaTyrGlyGlyTyrValPheIleLeuLeuIleProLeuHisValPheValLeu 280
808 TCTGCTGGGTGGGTGATGATTTATCATCAATCTTATCCACATGATGATTTGCTGTTG 867
281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
868 TTACTGATGCAGAGATACAGCAAGAGCTCTACATAGCATATAGCACTTTCTACATTGTG 927
301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
928 GGTTTAATATATCAATGATGATACCTTTTGGGATTTCCAGCAATCAGAACAGTGAA 987
321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
988 CACATGGCAGCTGCAGGTGCTTTGCAATTTGCTGCAAGCTTATGCTTTCTGCAATCTG 1047
341 ArgAspArgLeuThrIleGlnGlnPheGlnThrLeuPheLeuGlyValSerLeuAla 360
1048 AGAGACCGGATTAACAAACAGAGTTCCAGACCTTTCTTTTGGGTGATCACTAGCT 1107
361 AlaGlyValAlaPheLeuSerValIleTyrLeuThrTyrIleAlaProTyr 380
1108 GCAGGTGCTGTTCTCTTAGTCTATCTATTGATCTTATACAGTTACATTTGACCATGG 1167
381 SerGlyArgPheTyrSerLeuThrPheThrGlyTyrAlaIleHisIleProIle 400
1168 AGTGGCAGGTTTATTCATTTGGGATGATGCGGATGCAAAATACATTTCCAATATT 1227
401 AlaSerValSerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIle 420
1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTCTTTCTTTGATCTACATATT 1287
421 LeuValCysThrPheProAlaGlyLeuTyrPheCysIleLeuAsnIleAsnAspGluArg 440
1288 CTGTATGATACCTTCCAGAGGCTTTGGTCTGATCAAAATATCAAGATGAAAGA 1347
441 ValPhe 442
1348 GTATTT 1353

ORIGIN

1 MetalGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
28 ATGGGGAGCCCTCGGCGCGGAGAGCAAGCAGTCTCTCACTCGTCCCGGTGG 87
21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
88 AGTGGCCTCATGGCCCTGGGAACAGCGGACGCGGACCCAGCGGCGGCGGCGGCGG 147
41 AlaHisLysAlaAlaGlyGlyAlaAlaProPheProAlaProAlaGlyLeuSerGly 60
148 GCGCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207
61 GlyLeuSerGlnProAlaGlyTyrGlnSerLeuSerPheThrIleLeuPheLeuAla 80
208 GGGCTGTGCGACCGCGGCTGGGTGCGAGTCTCTCTCTCACTACCATCTCTCTCTG 267
81 ThrLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
268 TGGCTTGGCGCTTCAGTCTCGGCTCTTCCGCTCATCGCTTCGAAAGCATCATCCAC 327
101 GluPheAspProTyrPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyr 120
328 GAGTTCGACCGGCTGTTTAACTATAGATCAACACATCATCTTGCATCTCATCGGTTCTAT 387
121 GluPheLeuAsnThrPheAspGluArgAlaTyrTyrProLeuGlyValGlyVal 140
388 GAATTTTAAATGTTTGAAGAGAGCATGATGATCCACTAGGAAGATAGTAGTGGT 447
141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisThrIleLeuAsnThrLeu 160
448 ACTGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCAATGGATTTTAAATACATG 507
161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
508 AACATACTGTTTACATAAGAGAGCTGTGTGTCTTCTTGCACCACTTTTAGCGGCTT 567
181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuThrPsnGlnGlyAlaGlyLeuLeu 200

Alignment Scores:

Pred. No.: 0 Length: 1664
Score: 442.00 Matches: 442
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.51% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x AK074587 (1-1664)

RESULT 7

AX799084 2710 bp mRNA linear PAT 08-OCT-2003
Sequence 3 from Patent WO03054008.
AX799084
AX799084.1 GI:37605059
Mus musculus (house mouse)
Mus musculus
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .2710

REFERENCE

AX799084 2710 bp mRNA linear PAT 08-OCT-2003
Sequence 3 from Patent WO03054008.
AX799084
AX799084.1 GI:37605059
Mus musculus (house mouse)
Mus musculus
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .2710

FEATURES

AX799084 2710 bp mRNA linear PAT 08-OCT-2003
Sequence 3 from Patent WO03054008.
AX799084
AX799084.1 GI:37605059
Mus musculus (house mouse)
Mus musculus
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .2710

Primer for synthesizing full-length cDNA and use thereof FH Key

| FT | CDS | Location/Qualifiers | (28)...(1359) |
|----------|--------------------------|-------------------------|-----------------------|
| FEATURES | Location/Qualifiers | 1..1664 | |
| source | /organism="Homo sapiens" | /mol_type="genomic DNA" | /db_xref="taxon:9606" |

Alignment Scores:

Pred. No.: 0 Length: 1664

Score: 442.00 Matches: 442

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 53.51% Indels: 0

DB: 6 Gaps: 0

ORIGIN

US-10-028-384-2 (1-826) x BD127193 (1-1664)

QY 1 MetAlaGluProSerAlaProGluSerHisLysSerSerLeuAenSerSerProTyr 20

Db 28 ATGGCGGAGCCCTCGGCCCGGAGAGCAAGTCGTCCTCAACTCGTCCCGGTG 87

QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40

Db 88 ATGTGCCTCATGGCCCTGGGAAACAGCGGACCGCCACCGGGCCGGGGCCAGTGC 147

QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60

Db 148 GGGACACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207

QY 61 GlyLeuSerGlnProAlaGlyTyrGlnSerLeuSerPheThrLeuPheLeuAla 80

Db 208 GGGCTGTGCGACCGCGCTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 267

QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100

Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTTCATCGCTTCGGAAGCATCCAC 327

QY 101 GluPheAspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120

Db 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTCATCTCATGGTCTAT 387

QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGly 140

Db 388 GAATTTTAAATGCTTTGATGAAAGACATGGTATCCACTAGGAAGATAGTAGTGT 447

QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160

Db 448 ACTGTTTACCCAGGCTTATGATACCGCTGCCTTATTCATTCGATTTTAAATACAT 507

QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180

Db 508 AACATAACTGTTTCACATAAGAGACGTGTGTCTTCCTTCACCAACTTTTAGCGCCT 567

QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200

Db 568 ACATCTATCTACTCTTCCTGCTTACAGAGAATCTTGGACCAAGGAGGAGGAGCTTTA 627

QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220

Db 628 GCTGCTGTTGTTTATTGCTATTGTTACAGGCTACATATCTCGTCAGTAGCTGATCT 687

QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIysSer 240

Db 688 GATATGAGGAGGATGCTATTTTTCGACTTCAGTTCACATCTATTTATGGTAAATCT 747

QY 241 ValIysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260

Db 748 GTAAACATGGCTGCTGCTTTTGGACATATGCTGCTTATCTCTATCTATATGCT 807

QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280

Db 808 TCTGCTTGGGCTGGTATGATTTATTCATCAATCTTTATTCACATGATTTTGTGTG 867

QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300

Db 868 TTACTGATGACAGATACAGCAAGAGTCTACATAGCATATAGCACATTTCTACATG 927

QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320

Db 928 GGTTTAAATATTCATATGTCAGATACCTTTTGGGATTCAGCCATCAGAACAGTGA 987

QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340

Db 988 CACATGCGAGCTGCGAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTCAGTATCT 1047

QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360

Db 1048 AGAGACCGATTAAACAAACAGAGTTCAGAGCCCTTTCTTTTGGGTGTATCATAGCT 1107

QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380

Db 1108 GCAGTGTCTGTGTTCTTGTAGTGTCACTATTGACATTATACAGGTTTACATTCACCATG 1167

QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaIleIleHisIleProIleIle 400

Db 1168 AGTGGCAGGTTTATTCATTTGCGGATCTGGGTATGCAAAATACACATTCATTTAT 1227

QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePhePheAspLeuHisIle 420

Db 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTCTTCTTCTTGTATCATATT 1287

QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleIleAsnIleAsnAspGluArg 440

Db 1288 CTGTATGATCTCTCCAGAGGCTTTGGTCTGTCATCAAAATATCAACATGAAGA 1347

QY 441 ValPhe 442

Db 1348 GTATTT 1353

RESULT 6

AK074587

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AK074587 1664 bp mRNA linear PRI 03-SEP-2002

Homo sapiens cDNA FLJ90106 fis, clone HEMBA1006430, weakly similar to Human putative transmembrane protein precursor (B5) mRNA.

AK074587

AK074587.1 GI:22760122

oligo capping; fis (full insert sequence).

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S., Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Kojima, S., Nagahara, K., Masuho, Y., Ono, T., Okano, K., Yoshikawa, Y., Acsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Ninomiya, K.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 1664)

Isogai, T. and Otsuki, T.

Direct Submission

Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

```

/codon_start=1
/protein_id="CAP85831.1"
/db_xref="GI:45502339"
/translation="VAPSPAPESKHSKSLNSPWSGLMAGNSRHGHGPGACAAHKA
AGGAAPKAPAGSGLSQFQWQSLLSFTILFLANLAGFSRFLPAVIRFESIIEF
DFWNYRSTHLAGHGYEFLNWFDERAMYPGLRGVTVYCLMGTAGLIHWIINTL
NTHIARDVCFAPLFTSGLTSTFLTRLENOAGLAACTAIAPGYISRVAG
SPDNIEGAI PALQFTYLYLWKVSKTGVFVMTCCLSYFVMSAWGVYFIINLPLH
FVLLMCRYSKRVYIAYSTFYVGLILSNQIPVGFQPIRSEHMAAGVALLQAY
AFLOYLRDLTKQRFQTLFPLGSLAGAVFLSVILYTGTHIAPNSGRFSLMDY
AKIHPIIASVSEHQPTTWSFFDLHLVCTPPAGLWFCIKMNSERVFGA"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 0 Length: 1664
Score: 442.00 Matches: 442
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 53.51% Indels: 0
DB: 6 Gaps: 0

```

US-10-028-384-2 (1-826) x CQ782554 (1-1664)

```

QY 1 MetAlaGluProSerAlaProGluSerIlyValHisSerSerLeuAsnSerSerProTTP 20
DB 28 ATGGGGAGCCCTGGGCCCCGGAGAGCAAGATCGTCCCTCAACTCGTCCCCGTGG 87
QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
DB 88 AGTGGCCCTCATGGCCCTGGGAAACAGCGCCGACCCACCGCGCGCCGCGCCAGTGC 147
QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProlySerProAlaProAlaGlyLeuSerGly 60
DB 148 GGGCCACAGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 207
QY 61 GlyLeuSerGlnProAlaGlyTTPGlnSerLeuSerPheThrIleLeuPheLeuAla 80
DB 208 GGGCTGTCCGACCGCGCTGGGTGGCGTGGCTGTCTCTCTCCATCCATCTCTCTCT 267
QY 81 TTPLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100
DB 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCCCGCTCATCCGCTTCGAGGATCATCCAC 327
QY 101 GluPheAspProTTPPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
DB 328 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTGCATCTCATGGGTCTAT 387
QY 121 GluPheLeuAsnTTPPheAspGluArgAlaTTPProLeuGlyArgIleValGlyGly 140
DB 398 GAATTTTAAATGGTTTGAATCAAGAGCATGTATCCATAGAGAGATAGTAGTGT 447
QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTTPLeuAsnThrLeu 160
DB 448 ACTGTTTACCCAGGCTTCATGATACCGCTGGCTTATTCATGGATTTTAAATACATG 507
QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
DB 508 ACATTAATGTTCATAGAGAGAGCTGTGTGTCTTCCACCACTTTTACGGCCCTT 567
QY 181 ThrSerIleSerThrPheLeuLeuThrArgGlnLeuTTPAsnGlnGlyAlaGlyLeu 200
DB 568 ACATCTATATCTCTCTTCTTACAGAGAGACTTTTGAACCAAGAGAGAGAGACTTTTA 627
QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
DB 628 GCTGCTGTTTATGTATGTATGTACAGGCTACATATCTGGTCAGTACGTGATCTTT 687
QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrIleTTPValLysSer 240
DB 698 GATAATGAAGCATTCGCTATTTTTCGCTTCAGTTCATCATCTATTTATGGGTAAATCT 747
QY 241 ValLysThrGlySerValPheTTPThrMetCysCysLeuSerTyrPheTyrMetVal 260

```

```

DB 748 GTAAACACGGTGGTCACTTTTGGCAAAATGCTGCTGCTATATCCATTTCTATATGCTC 807
QY 261 SerAlaTTPGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
DB 808 TCTGCTTGGGTGGTGTATGATTTATCATCAATTTATTCACATGCTCATGTTATTTGTGTG 867
QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
DB 868 TTACTGATGCAGAGATACAGCAAGAGCTACATAGCATATAGCACCTTTCTACATTTG 927
QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGln 320
DB 928 GGTTTAATATTAATCAATGCAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGA 987
QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
DB 988 CACATGGCAGCTGCAGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTCAGTATCTG 1047
QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
DB 1048 AGAGACCGATTAAACAAACAAAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1107
QY 361 AlaGlyValAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTTP 380
DB 1108 GCAGGTGCTGCTGCTTCTTAGTGTATCTATTGACTTATACAGGTTTACATTCACCATG 1167
QY 381 SerGlyArgPheTyrSerLeuTTPAspThrGlyTyrAlaLysIleHisIleProIleIle 400
DB 1168 AGTGGCAGGTTTTTATTCATTTGGGATATCGGTATGCAAAATAACATCATTCATTAAT 1227
QY 401 AlaSerValSerGluHisGlnProThrTTPValSerPhePhePheAspLeuHisIle 420
DB 1228 GCATCAGTGTCTGAGGATCAACTACGCTTGGGTGCTTTCTTCTTGTATCATATATT 1287
QY 421 LeuValCysThrPheProAlaGlyLeuTTPPheCysIleLysAsnIleAsnAspGluArg 440
DB 1288 CTTGTATGATCTCCACGAGCGCTTTGGTCTCTGTCATCAAAATAATCAACGATGAAGA 1347
QY 441 ValPhe 442
DB 1348 GTATTT 1353

```

RESULT 5
BD127193
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BD127193.1 GI:23222138
JP 2002017375-A/2624.
Homo sapiens (human)

1664 bp DNA linear PAT 18-SEP-2002
Primer for synthesizing full-length cDNA and use thereof.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 1664)
Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 2624 22-JAN-2002;
HELEX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2624
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII
PI YUKI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINTCHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
'10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC

QY 195 GInGlyAlaGlyLeuLeuAlaAlaCyaspheileAlaileValProGlyTyrIleSerArg 214
 DB 939 CAAGAGGAGGAGGACTTCTAGCTGCTTCTTCTATCTGATCGTACAGGAGTACATATCTCG 998
 QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlailePheAlaLeuGlnPheThr 234
 DB 999 TCAGTGGGGGATCCTTTGATAAGAGGCAATCCCATTTTGGCGCTTCACTTAC 1058
 QY 235 TyrLeuTyrValIleSerValIleThrGlySerValPheThrMetCysCysLeu 254
 DB 1059 TACTTATGGTAAAGTCTGTGAGACCGGCTCTGTCTTGGACAAATGCTGCTCTG 1118
 QY 255 SerTyrPheTyrMetValSerAlaTyrPheGlyTyrValPheIleIleAsnLeuIlePro 274
 DB 1119 TCATATTTCTCATGCTCTCTGCGTGGGAGGTTATGTGTTCATCAACCTCATCCCT 1178
 QY 275 LeuHisValPheValLeuLeuMetGlnArgTyrSerIleValPheAlaTyr 294
 DB 1179 CTCATGTGTGTCTGCTGCTGATGCGAGAGTACAGCAAGAGTCTACATAGCATAT 1238
 QY 295 SerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGln 314
 DB 1239 AGCATCTTCTCATTTGGGTTTAAATATTCATGCGATACCTTTTGTGGGATTCAG 1298
 QY 315 ProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyr 334
 DB 1299 CCAATCAGACAGGAGGACATGCGAGCTGACAGTGTCTTTGGCTGCTGCAAGCTTAC 1358
 QY 335 AlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnGluPheGlnThrLeuPhePhe 354
 DB 1359 GCTTTTGTGAGTATCTGAGAGACGGTGTGACAAACAGGAGTTCACAGACCTTTTCTTT 1418
 QY 355 LeuGlyValSerLeuAlaAlaGlyValPheLeuSerValIleTyrLeuThrTyrThr 374
 DB 1419 TTGGTGTCTCAGTCTGAGGCGCTGTGTTCTTAGTGTCTATCTATCTGACATACAC 1478
 QY 375 GlyTyrIleAlaProTyrSerGlyArgPheTyrSerLeuThrPheThrGlyTyrAlaLys 394
 DB 1479 GGTATATGACACATGAGTGGCAGGTTTATTCATCTATGGGATATCGGTTATGCAAAA 1538
 QY 395 IleHisIleProIleAlaSerValSerGluHisGlnProThrThrTyrValSerPhe 414
 DB 1539 ATACACATTCATATTTGTCATGATGTCGACATCAGCTCAGCATGGGTGCTTTC 1598
 QY 415 PhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuThrPheCysIleLys 434
 DB 1599 TTCCTTTGATCTACATATTTCTGTATGATACCTTCCAGCAGGCTATGGTTCTGCATCAA 1658
 QY 435 AsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAla 454
 DB 1659 AATAACAAGATGAAGAGTATTTGTGCTCTGTATGCGATCAGTGTCTACTTTGCC 1718
 QY 455 GlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIle 474
 DB 1719 GGAGTGATGGTGGCGCTGATGCTGACTCTGACCCCGCTGCTGCTGCTGCTGCTGCTG 1778
 QY 475 AlaPheSerAsnValPheGluHisTyrLeuGlyAspMetIleValArgGluAsnProPro 494
 DB 1779 GCCTTCTCAATGTTTTTGGACACTATTTGGGGATGACATGAAGAGGAAACCCACCT 1838
 QY 495 ValGluAspSerAspGluAspIleValArgGlnGlnGlyAsnLeuTyrAspLysAl 514
 DB 1839 GTGGAGGACAGCAGTGTGAGGATGACAAAGAACCC-AGGAACCTTGTATGACAGGC 1897
 QY 514 aGlyLysValArgLysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPro 533
 DB 1898 AGGTAAGTGAGGAAGCATGT-GACAGAGCAAGAA-ACCTGAAGAGGCTTGGGCCCC 1955
 QY 534 AsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetPheAlaValHis 553
 DB 1956 AACATCAAAAGCATTTGACCATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015

QY 554 CythrThrValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsn 573
 DB 2016 TGCACGTGGGTGCACAGCAACGCTACTCCAGTCCCAAGTGTGGTCTTGGCTCTACCAT 2075
 QY 574 HisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheThrPheLeuArgGln 593
 DB 2076 CATGATGGTACCAAGGATATATTAGATGATTTTAGAAGAACGCTACTTTTGGCTGAGACAA 2135
 QY 594 AsnThrAspGluHisAlaArgValMetSerTyrThrAspTyrGlyTyrGlnIleAlaGly 613
 DB 2136 AACACGGATGAACACCGCCCGGTCTGTGTGGTGGACCTACGGCTATCAGATTCTGGC 2195
 QY 614 MetAlaAsnArgThrThrLeuValAspAsnAsnThrThrAsnAsnSerHisIleAlaLeu 633
 DB 2196 ATGCCCAACAGGACCACTCTGGTGGATAACAACACCTGGAAACAACAGCCACATCCACTG 2255
 QY 634 ValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuAs 653
 DB 2256 GTCGNAAGCTATGTCTTCCATGAACCGCCCTATAAATCATGAGTCT-CCTTGA 2314
 QY 653 PValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAs 673
 DB 2315 TGTCCGATTTATGTGTGTGTTATTTCCGAGGAGTATGCTATTCGGGGAGCATATCAA 2374
 QY 673 nLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSe 693
 DB 2375 CAAGTCTCTGATGCTCAGGATAGCTGAAGGGGAGCATCCCAAAGACATCCGGGAAGG 2434
 QY 693 r-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLeu 712
 DB 2435 -TGACTATTTTCC-CCAGCAGGAGAGTTCGAGTAGACAAAGCTGGGTCTCTCTACTCTG 2492
 QY 713 LeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPhe 732
 DB 2493 TTAAACTGCCTTATGTATATAAATGTCTACTACAGATTTGGAGAAATGAGCTAGATTTT 2552
 QY 733 ArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLys 752
 DB 2553 CGCACTCCCCAGGCTTTGACCAACACGTAATGCTGAGATTGGAAATTAAGACATTAAA 2612
 QY 753 PheLysHisLeuGluGluAlaPheThrSerGluHisThrLeuValArgIleTyrLysVal 772
 DB 2613 TTCAAGCATTTGGAGGAAGCTTTTACATCAGAGCACTGGCTTGTGAGGATATATAAGTG 2672
 QY 773 LysAlaProAspAsnArgGluThrLeu 781
 DB 2673 AAAGCACTGACAAACAGGAGACACTA 2699

RESULT 4
 LOCUS CQ782554 1664 bp DNA linear PAT 17-MAR-2004
 DEFINITION Sequence 2694 from Patent EP1396543.
 ACCESSION CQ782554
 VERSION CQ782554.1 GI:45502338
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 TITLE Primers for synthesizing full length cDNA clones and their use
 JOURNAL Patent: EP 1396543-A 2694 10-MAR-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 source 1..1664
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 28..1362
 /note="unnamed protein product"

CDS


```

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
BC052433.1 GI:30851501
MGC.
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
1 (bases 1 to 4236)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Schett,T.B., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahay,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Schmutz,J.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywicki,M.I., Skalska,U., Smallos,D.E.,
Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 4236)
Strausberg,R.
Direct Submission
Submitted (15-MAY-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: http://genome.uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldo,M.F., Akabogu,I., Bair,T., Crouch,K., Davis,A.,
Fishler,K., Keppel,C., Kucaba,T., Lebeck,M., Melo,A., Schaefer,K.,
Scheetz,T., Smith,C., Snitz,E., Tack,D., Trout,K., Walters,J.,
Casavant,T., Soares,M.B.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: Place: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
Location/Qualifiers
1. 4236
/organism="Mus musculus"
/mol_type="cDNA"
/db_xref="taxon:10090"
/clone="MGC:64679 IMAGE:6837097"
/tissue_type="Brain, enriched mouse brain 12.5dpc"
/clone_lib="NIH BMAP_F10"
/lab_host="DH10B"
/notes="Vector: pYX-ASC"
1. 4236
/gene="1300006C19rik"
/notes="synonym: Simp"
/db_xref="LocusID:68292"
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 4236
Score: 452.00 Matches: 721
Percent Similarity: 98.36% Conservative: 0
Best Local Similarity: 98.36% Mismatches: 6
Query Match: 54.72% Indels: 12
DB: 10 Gaps: 0
US-10-028-384-2 (1-926) x BC052433 (1-4236)
Qy 55 ProAlaGlyLeuSerGlyLeuSerGlnProAlaGlyTyrGlnSerLeuSerPhe 74
Db 519 CCCGCGGGGCTGTCCGGGGCTGTGCGACGCGCGGGTGGCAGTCTGCTCTCTCCTTC 578
Qy 75 ThrIleLeuPheLeuAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIleArg 94
Db 579 ACCATCTCTTCTGCGCTTGGCTGGCGGCTTCAGCTCGCGCTTCTGCGGTATCCGC 638
Qy 95 PheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgSerThrHisLeu 114
Db 639 TTCGAGAGCATCATCCACAGTTCGACCGCGTGGTTAACTATAGATCAACACATCATCT 698
Qy 115 AlaserHisGlyPheTyrGluPheLeuAsnTyrPheAspGluArgAlaTyrTyrProleu 134
Db 699 GCATCTCATGGATCTATGAGTTTCTAAATGGTTGATGAAGAGCATGTTACCCACTG 758
Qy 135 GlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHis 154
Db 759 GGAAGAATAAGTGGGTGGCCCGCTTACCAGAGGTTGATGATAACAGCTGGGCTTATTTCAT 818
Qy 155 TrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAla 174
Db 819 TGGATTTTAAATACATTGAACATACAGTTTCACATAAGAGATGTGTGTATTCTCTTGA 878
Qy 175 ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTyrAsn 194
Db 879 CCAACTTTTAGCGGCGCTTACATCCATATCTACGTTCTCTGCTAACTAGAGAACTGTGGAC 938

```

141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisThrPheLeuAsnThrLeu 160
142 ACTGTTTATCCAGGGTGTATGATAACCGCTGGCTTATTCATTGGATTTTAAATACATATG 480
161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
162 AACATAACTGTTTACATAAGAGAGCTATGTGTGTTTCCCTTGCACCAACTTTTAGCGGCTT 540
181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuThrPheAsnGlnGlyAlaGlyLeuLeu 200
182 ACATCTATATCTACTTCTCTGCTTACAAGAGAACTTTGGAACCAAGAGAGAGAGACTTTTA 600
201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
202 GCTGCTGTGTTTATTGCTATTGTACAGGCTACATATCTCGTTCAGTAGCTGGATCCTTT 660
221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuThrValIleSer 240
222 GATAATGAAGGCAATGCTATTTTGCACCTTCAGTTCACATACATATTTATGGTAAATCT 720
241 ValIleThrGlySerValPheThrMetCysCysLeuSerTyrPheTyrMetVal 260
242 GTAAAAAATCGGCTCAGTCTTTTGGACAATGTGCTGCTTATCTTATCTATATGCTC 780
261 SerAlaThrGlyGlyValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
262 TCTGCTGGGTGGTATGATATTTATCATCATCTTATCCACTGCAATGATTTTGTGTTG 840
281 LeuLeuMetGlnArgTyrSerIleValTyrIleAlaTyrSerThrPheTyrIleVal 300
282 TTACTGATGCAGATACAGCAAAAGAGCTACATAGCATATAGCATTTTCTACATGTTG 900
301 GlyLeuIleLeuSerMetClnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
302 GGTTHATATATCAATGAGATACCTTTTGGGATTTCCAGCCATTCAGAAACAGTGA 960
321 HisMetAlaAlaGlyValPheAlaLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
322 CATATGGCAGCTGCAGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTGGAGTATCTG 1020
341 ArgAspArgLeuThrIleGlnIlePheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
1021 AGAGACCGATTAACAAACAGAGTTCAGACCTTTTCTTTTGGGTGATCAGTACCT 1080
361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380
1081 GCAGGTGCTGCTTCTTATGATCATCTATTGCTATATGCTTATACAGGTTACATGCAATGG 1140
381 SerGlyArgPheTyrSerLeuThrPheThrGlyTyrAlaIleHisIleProIleIle 400
1141 AGTGGCAGGTTTATTCATTGTTGGATCTGGGTATGCAAAATACACATTCACATATAT 1200
401 AlaSerValSerGluHisGlnProThrThrTyrValSerPhePheAspLeuHisIle 420
1201 GCATCAGTGTCTGAGCATCACTACGACTTGGGTGCTTCTTCTTGTGATCTACATAT 1260
421 LeuValCysThrPheProAlaGlyLeuThrPheCysIleIleValAsnIleAsnAspGluArg 440
1261 CTGTATGATPACCTTCCAGCAGGCTTTGGTTCGTATCAAAAATATCAAGATGAAGA 1320
441 ValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeu 460
1321 GTATTGTTGCTCTATATGCAATCAGTCTGCTTCTTCTTGTGAGTGTGCTGCTG 1380
461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
1381 ATGTTGACTTTTGCCTCCAGTCTGTGTATGCTGCTGCAATGCTTTTCAATATGTTTT 1440
481 GluHisTyrLeuGlyAspMetIleArgGluAsnProProValGluAspSerSerAsp 500
1441 GAGCACTATTGTTGGGGATGACATGAAAAGGAAAAATCCACTGTGGAGGAGCAGCAGTAT 1500
501 GluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHis 520

1501 GAGGATGCAAAAAGAAACCAAGGAAATTTCTATGATAAGCGAGGTAAATGTGAGGAAACAT 1560
521 AlaThrGluGlnGluIleThrGluGlyLeuGlyProAsnIleLysSerIleValThr 540
1561 GCAACTGAACAGGAAAAAACTGAAGAGGATTTAGGCCCTTAATATAAAAAGCATTTGTCCAC 1620
541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrThrValThrSerAsn 560
1621 ATGTTGATGCTGATGCTTATTTGCTGCTACCTGCTACCTGGGTCAACAGCAAT 1680
561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
1681 GCCTACTCTAGTCCAAAGTGTAGTCTGCTGCTCATCAATCATGATGCCACCAAGGATATC 1740
581 LeuAspAspPheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGluHisAlaArg 600
1741 TTAGATGATTTTAGAAGGCTTTACTTTTGGCTAAGCAAAATACAGATGAACATGACGCA 1800
601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
1801 GTAAATGCTTGGTGGGATTTATGCTATCAGATAGCTGGAATGCTTAATAGAACTACGTTG 1860
621 ValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
1861 GTGATATATACACTGCAATTAACAGCCACATAGCACTGTGGGAAAAGCTATGCTTCT 1920
641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
1921 AATGAAACAGCAGCCTATAAAATCATGAGGACTCTAGATGTAGATTTATGTTGTTAT 1980
661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnIlePheLeuThrMetValArg 680
1981 TTTGGAGGGTATTGCTTATTTCTGGGATGATATCACAAAATTTCTCTGGATGTTAG 2040
681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
2041 ATAGCTGAAGAGAGAAATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGGA 2100
701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
2101 GAATTCGTTGATCAAAAGCAGGATCCCTTCTTGTGATTTGCTTATGATATAAATG 2160
721 SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
2161 TCATACATACATTTGGAGAAATGCACTGGATTTTCTGACACCCCAAGCTTTGACCGA 2220
741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisIleGluGluAlaPhe 760
2221 ACAGTAAATGCTGAGATTTGAAATTAAGACATTTAAATTCACATTTGGAAGAGCTTT 2280
761 ThrSerGluHisThrLeuValAlaGlyTyrLysValLysAlaProAspAsnArgGluThr 780
2281 ACATCAGAACACTGGCTTGTAGGATATAATAAGTAAAGACCTGATTAACAGGAGACA 2340
781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
2341 TTAGATCAAAACCTTCAGTCAACACATTTTCCCAAAACAGAGATTTTGTTCACAGAG 2400
801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
2401 ACTACAAAAGAGAGCGTGGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAA 2460
821 IleSerLysLysThrVal 826
2461 ATATCTAAGAGAGACTGTT 2478

RESULT 3

BC052433

LOCUS

DEFINITION Mus musculus RIKEN cDNA 1300006C19 gene, mRNA (cdna clone MGC:64679 IMAGE:6837097), complete cds.

ACCESSION BC052433

4236 bp

linear

ROD 12-NOV-2003

QY 601 ValMetSerTrpTyrPheGlyTyrGlnIleAlaGlyMetAlaSerArgThrLeu 620
 DB 1801 GTAATGCTTGTGGGATATGCTATCAGATAGCTGGAATGGCTAATAGACTAGCTTG 1860
 QY 621 ValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIleValSerSer 640
 DB 1861 GTGGATATATACCTGGAATACAGCCATAGACTGCTGGGAAAGCTATGCTTT 1920
 QY 641 AsnGluThrAlaIleAlaTyrIleMetArgThrLeuAspValAspTyrValLeuValIle 660
 DB 1921 AATGAACAGCAGCCTATATAAATCATGAGACTCATAGATGATATGTTTGGTTATT 1980
 QY 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
 DB 1981 TTGGAGGGTATTGGCTATCTGCTATGCTGATGATATCAACAAATTCCTCGATGGTTAGG 2040
 QY 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 DB 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCACAGGGA 2100
 QY 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 DB 2101 GAATTCCTGCTAGACAAAGAGAGATCCCTACTTTGTTGATTCCTTATGATATAAATG 2160
 QY 721 SerTyrTrpArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArg 740
 DB 2161 TCATACATCAGATTTGGAGAAATGCGCTGGATTTTCGTACACCCCGAGTTTGTACCGA 2220
 QY 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 DB 2221 ACAGTAAATCTGAGATTTGGAAATAGGACATTAATTAATTAATTAATTAATTAATTAAT 2280
 QY 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 DB 2281 ACATCAGAACACTGCTGTTAGGATATATAAGTAAAGCACCTGATACAGGAGACA 2340
 QY 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 DB 2341 TTATGATACAAACCTCGATGATACCAACATTTCCCAAAACAGAGATTTTGTCAAGAAG 2400
 QY 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 DB 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAAGCTGTTTAAAGAAAGCAAGAA 2460
 QY 821 IleSerLysLysThrVal 826
 DB 2461 ATATCTAAGAGACGTGTT 2478
 RESULT 2
 AY074880
 LOCUS Homo sapiens source of immunodominant MHC-associated peptides
 DEFINITION (SIMP) mRNA, complete cds.
 ACCESSION AY074880
 VERSION AY074880.1 GI:19879588
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2481)
 Mc Bride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A.,
 Bergeron, J., and Perreault, C.
 The model B2dmi minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast SPT3 gene
 Immunogenetics 54 (8), 562-569 (2002)
 23326278
 PUBMED 12439619
 2 (bases 1 to 2481)
 Mc Bride, K. and Perreault, C.
 Direct Submission
 TITLE Submitted (22-JAN-2002) Molecular Biology Group, Compaticigene, 6100
 JOURNAL Royalmount, Montreal, Qc H4P 2R2, Canada

FEATURES
source

gene

CDS

Location/Qualifiers
 1. 2481
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 1. 2481
 /gene="SIMP"
 /gene="SIMP"
 /notes="similar to Saccharomyces cerevisiae OSTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"
 /protein_id="AAL71884.1"
 /db_xref="GI:19879588"
 /translation="MAEPSAPESKHSLSNPSWGLVALNSRHGHGPGACAHKA
 AGAAPKAPAGLSGQAGWSLISFTILFLAWLAFSSRLFAVIRFESIHEF
 DPWFVRSVTHLASHGFEFLNWFDERAWYPLGRVGTVPGLMTVGLHILNLT
 NITVTRDVCVFLPTFSLTSTISFTLRELNOGAGLAACTAIIVGYSRSVAG
 SPNEGIALEALQFYLYLVKSVKTSVFVWMCCLSVFYVSVAGGVFIINLPLH
 VFVLLMQRYSKRVIAVSTFYVGLIILSMQIPFVGFQPIRTSEHMAAGVFAIQAY
 AFLQURDLTKQEFQTLFFLGLVLAAGAVFLSVIYLYTGTIAPWSGRFVSLNDTGY
 AKTHPIIASVSEHPTTVSFDFDLHLVCTFPAGLWFCIKNINDESVFVALYISA
 VYFAGVYVRLMLTLTPVVCMLSAIAFNVFHYLGGDDMKRENPPVEDSDDKRNQ
 NYDRAKVRKHATEQTEEGIPNKSIVTMLMLLMAFVHTVWTNSAYSSPS
 VVLASVNDGTRNLDIFREAVFLNRQITDEHARVMSWMDYGYQTAGMANETTLVDNN
 TWNSHIALVGRAMSNETAAYKIMRTLDVDVLYFGVIGYGGDDINKELMWVRIA
 EGHPKDJESDYFTPOGEFRVDKAGSTLLNCLMVKSYRFGEMQDLPRTPGFDR
 TRNAELGNDIKFKLEAFTSEHNLVLYKVKAPDNRETLIDHPRVTNIPFKQKYL
 KKTTKRGRGYIKNKLVEKKKIKSKTV"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2481
 Score: 826.00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-2 (1-826) x AY074880 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 DB 1 ATGGCGAGCCCTCGGCCCGGAGACAGACAGTCTCTCACTCGTCCCGCTGG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCTCATGCGCCCTGGGAAACAGCGGACGCGCACACGCGGCCCGCGGCGCCAGTGC 120
 QY 41 AlaHisLysAlaAlaGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 CGCACAGCG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGGCTGTGCGAGCGCGGTGGTGGAGTCTCTCTCTCTCCATCTCTCTCTCTCTCTCTCT 240
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 DB 241 TGGCTTCGCGGCTTACCTCGCCCTCTTCGCGGCTCATCGCTTCGAAAGCATCATCCAC 300
 QY 101 GluPheAspProTrpPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
 DB 301 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTTCGATCTCATCGGTTCTAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluAlaTyrTrpProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATTTGGTTTGTATGAAGAGCATGGTATCCACTAGGAAGAAATAGTAGTGT 420

AGAAAPPPAPAGLSGGLSQPAGWQSLSTFLFLAWLAGFSRLFAVIRPFSITHF
 DWFNRYTHSLASGFGFVFNWDERAWFLRGVGTVYGLMIGTALHWILNLT
 NINITHIRVDFVLAFTPSGLSTSLFLTRFLNGLWGLAGAACFAIVFGYISVSG
 SPDNFENALFALQYVILASGAVKTSVFTWCCVGSFVYMSAAWGAAGVILNPLH
 VFFLLMORYSRTYVIAVSTVYVGLILSMQIPFGFQFPIRSEHMAAGVAFVALLQ
 AFLOYLRDLTKQESFTFLFGLASAGVFLSVILYTGTVLAWSGFVSLWDITGY
 AKIPIHITASVHEQPTWVPSFEDHITLQTPAGVWFCQIKNIDRBFVFLYALSA
 VYVFGVWFLMLTFLVCMUSALFASVNFVHGLDDMKZENPVEDSDEDDKQOG
 NVLYDKAGVRKHATCEKTEPGLGNQSTLMLMLMMFVAFQIAGMARTVTSNAYSFS
 NVLSVYNGHDNTLIDPREAYFMLRQNTDIEHARMSWMDGYQIAGMARTVTDVNN
 TWNGSHALIVAGMSNASTKYINRDLIDVLYFGVGYSGSDINKFLMPVIRFA
 EGEHGHQTRADSESTVYFQGEFRVDRKAGSDPTLNLICNGLMSYVRGEGQDLPFPFDR
 TNAEIGNKDIKFHELEASKEHSHVTVYVKVAPDNRELTDHKPRVTWVIFPKQKYS
 KTKTKRGGCVIKNKIVFKKKKISKVLT"

ORIGIN

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 0 | Length: | 2481 |
| Score: | 826.00 | Matches: | 836 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DS: | 6 | Gaps: | 0 |

US-10-028-384-2 (1-826) x AX799082 (1-2481)

| | | |
|-----|---|-----|
| 1 | MetAlaGluProSerAlaProGluSerLyshisYserSerLeuAnSerSerProTrp | 20 |
| Qy | | |
| 1 | ATGGCGAGCCCTCGGCCCGGAGAGCAAGTCGCTCAACTCGTCCCGCGTG | 60 |
| Db | | |
| 21 | SerGlyLeuMetAlaLeuGlyVasnSerArgHisGlyHisGlyProGlyAlaGlnCys | 40 |
| Qy | | |
| 61 | AGTGGCCTCATGGCCCTGGGAAACACAGCCGCACCGCCACACAGGGCCCGGCGCCAGTGC | 120 |
| Db | | |
| 41 | AlaHisLyAlaAlaGlyGlyAlaAlaProProLySerProAlaProAlaGlyLeuSerGly | 60 |
| Qy | | |
| 121 | CGCCACAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGTCCCGG | 180 |
| Db | | |
| 61 | GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrLeuLeuPheLeuAla | 80 |
| Qy | | |
| 181 | GGGTGTTCGCAGCCGGGTGGGTGGACGTGCGCTCTCTCTTCACCATCTCTTCTGTGCC | 240 |
| Db | | |
| 81 | TrpLeuAlaGlyPheSerSerArgLeuPheAlaValleArgPheGluSerIleIleHis | 100 |
| Qy | | |
| 241 | TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGGTCTCGAAGCATCATCCAC | 300 |
| Db | | |
| 101 | GluPheAspProTrpPheAsnTyArgSerThrHisLeuAlaSerHisGlyPheTyTr | 120 |
| Qy | | |
| 301 | GAGTTCGACCCGCTGGTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCAT | 360 |
| Db | | |
| 121 | GluPheLeuAsnTrpPheAspGluArgAlaTrpTyProLeuGlyArgIleValGlyGly | 140 |
| Qy | | |
| 361 | GAATTTTAAATTGGTTTGATGAAAGAGCATGTGTATCCACTAGGAAGAATAGTAGTGGT | 420 |
| Db | | |
| 141 | ThrValTyProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAnThrLeu | 160 |
| Qy | | |
| 421 | ACTGTTTACCAGGGTTGATGATAAACGCTTGGCGCTTATTCATTGGATTTTAAATCATGT | 480 |
| Db | | |
| 161 | AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu | 180 |
| Qy | | |
| 481 | AACATAACTGTTCATAGAAGAGCGTATGTGTGTTCTTGCACCAACTTTTAGCGCGCTT | 540 |
| Db | | |
| 181 | ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu | 200 |
| Qy | | |
| 541 | ACATCTATATCTACTTCTCTCTTACAAGAGACTTTGGAACCAAGGACGAGACTTTTA | 600 |
| Db | | |
| 201 | AlaAlaCysPheIleAlaIleValProGlyTyIleSerArgSerValAlaGlySerPhe | 220 |
| Qy | | |
| 601 | GCTGTCTGTTTTATTGCTATTGTACCGAGGCTACATATCTCGGTACGTAGTACCTTT | 660 |
| Db | | |
| 221 | AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyTrLeuTrpVallysser | 240 |
| Qy | | |
| 661 | GATATGAAGCATGTATTATTGGCACTTCAGTTACATCATATTATTAGGGTAAAATCT | 720 |
| Db | | |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57; Search time 9151.03 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPSKHKSINSSPW.....GYKNKLKLVFKGKKISKKTIV 826

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-O=/csm2.1/USP20 spool/US10028384/runat_14122004_131645_6800/app_query.fasta.1.3740
-DB=GenEmbl -QMT=fscap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.1 20262 @runat_14122004_131645_6800 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.on.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 826 | 100.0 | 2481 | 6 | AX799082 |
| 2 | 826 | 100.0 | 2481 | 9 | AY074880 |
| 3 | 452 | 54.7 | 4236 | 10 | BC052433 |
| 4 | 442 | 53.5 | 1664 | 6 | CQ782554 |

| | | | | | | |
|----|-----|------|--------|----|----------|--------------------|
| 5 | 442 | 53.5 | 1664 | 6 | BD127193 | BD127193 Primer fo |
| 6 | 442 | 53.5 | 1664 | 9 | AK074587 | AK074587 Homo sapi |
| 7 | 365 | 44.2 | 2710 | 6 | AX799084 | AX799084 Sequence |
| 8 | 323 | 39.1 | 2546 | 6 | AX099510 | AX099510 Sequence |
| 9 | 323 | 39.1 | 2546 | 6 | BD063986 | BD063986 Secreted |
| 10 | 317 | 38.4 | 2510 | 6 | AX136135 | AX136135 Sequence |
| 11 | 317 | 38.4 | 2510 | 6 | BD123520 | BD123520 Secretary |
| 12 | 317 | 38.4 | 2510 | 9 | AK075380 | AK075380 Homo sapi |
| 13 | 289 | 35.0 | 2503 | 6 | AX017997 | AX017997 Sequence |
| 14 | 289 | 35.0 | 2503 | 6 | BD137303 | BD137303 Human nuc |
| 15 | 286 | 34.6 | 2284 | 6 | AX882932 | AX882932 Sequence |
| 16 | 286 | 34.6 | 2284 | 6 | BD160013 | BD160013 Primer fo |
| 17 | 286 | 34.6 | 2284 | 9 | AK027789 | AK027789 Homo sapi |
| 18 | 246 | 29.8 | 957 | 9 | BC015880 | BC015880 Homo sapi |
| 19 | 195 | 23.6 | 1734 | 10 | BC003206 | BC003206 Mus muscu |
| 20 | 168 | 20.3 | 507 | 6 | AX886102 | AX886102 Sequence |
| 21 | 168 | 20.3 | 507 | 6 | BD025712 | BD025712 Sequence |
| 22 | 167 | 20.2 | 787 | 6 | AX869456 | AX869456 Sequence |
| 23 | 167 | 20.2 | 787 | 6 | BD149518 | BD149518 Primer fo |
| 24 | 167 | 20.2 | 1040 | 6 | CQ723424 | CQ723424 Sequence |
| 25 | 145 | 17.6 | 784 | 6 | AX136480 | AX136480 Sequence |
| 26 | 145 | 17.6 | 784 | 6 | BD123720 | BD123720 Secretary |
| 27 | 143 | 17.3 | 433 | 6 | AX340470 | AX340470 Sequence |
| 28 | 133 | 16.1 | 558 | 6 | AX079689 | AX079689 Sequence |
| 29 | 123 | 14.9 | 2508 | 10 | BC013054 | BC013054 Mus muscu |
| 30 | 117 | 14.2 | 616 | 6 | BD210442 | BD210442 Human gen |
| 31 | 114 | 13.8 | 2620 | 5 | CR386955 | CR386955 Gallus ga |
| 32 | 107 | 13.0 | 349 | 6 | BD071886 | BD071886 Secreted |
| 33 | 105 | 12.7 | 170691 | 9 | AC104643 | AC104643 Homo sapi |
| 34 | 97 | 11.7 | 101046 | 2 | AC117975 | AC117975 Rattus no |
| 35 | 97 | 11.7 | 203492 | 2 | AC107826 | AC107826 Mus muscu |
| 36 | 97 | 11.7 | 227953 | 2 | AC134292 | AC134292 Rattus no |
| 37 | 97 | 11.7 | 285379 | 2 | AC113070 | AC113070 Mus muscu |
| 38 | 82 | 9.9 | 250 | 6 | AX885575 | AX885575 Sequence |
| 39 | 82 | 9.9 | 250 | 6 | BD025185 | BD025185 Sequence |
| 40 | 81 | 9.8 | 49764 | 2 | AC101107 | AC101107 Mus muscu |
| 41 | 78 | 9.4 | 245 | 6 | AX905227 | AX905227 Sequence |
| 42 | 78 | 9.4 | 245 | 6 | BD040760 | BD040760 Sequence |
| 43 | 78 | 9.4 | 247 | 6 | BD075907 | BD075907 5' EST of |
| 44 | 72 | 8.7 | 177253 | 9 | AC092024 | AC092024 Homo sapi |
| 45 | 63 | 7.6 | 1748 | 5 | BC052114 | BC052114 Danio rer |

ALIGNMENTS

| | | | | | | |
|------------|--|--|---------|------|--------|-----------------|
| RESULT 1 | AX799082 | Sequence 1 from Patent WO03054008. | 2481 bp | mRNA | linear | PAT 08-OCT-2003 |
| LOCUS | AX799082 | | | | | |
| DEFINITION | AX799082 | | | | | |
| ACCESSION | AX799082 | | | | | |
| VERSION | AX799082.1 | GI:37605057 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | | | | | | |
| ORGANISM | | | | | | |
| | Homo sapiens | | | | | |
| | Homo sapiens | | | | | |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | | |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| REFERENCE | | | | | | |
| | 1 | Perreault, C. and McBride, K. | | | | |
| | | Mammalian SAMP protein, gene sequence and uses thereof in cancer | | | | |
| | | therapy | | | | |
| AUTHORS | | | | | | |
| TITLE | | | | | | |
| JOURNAL | | | | | | |
| | Patent: WO 03054008-A 1 03-JUL-2003; | | | | | |
| | Compatisigne Inc. (CA) | | | | | |
| FEATURES | | | | | | |
| | Location/Qualifiers | | | | | |
| source | | | | | | |
| | 1..2481 | | | | | |
| | /organism="Homo sapiens" | | | | | |
| | /mol_type="mRNA" | | | | | |
| | /db_xref="taxon:9606" | | | | | |
| | 1..2481 | | | | | |
| | /codon_start=1 | | | | | |
| | /product="SAMP" | | | | | |
| | /protein_id="CAE48447.1" | | | | | |
| | /db_xref="GI:37605058" | | | | | |
| | /translation="MAEPSAPSKHKSINSSPWKSLMALGNRSHGHGPGCAQCAHKA" | | | | | |

Result No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```
Db 303 AATGAAACAGCAGCCTATATAAATCATGAGGACTCTAGATGTAGATTATGTTGGTTATT 362
Qy PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArg 680
Db TTTGGAGGGTTATTGGCTATTCTGGTGATGATATCAACAATTTCTCTGGATGGTTAGG 422
Qy IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
Db ATAGCTGAAGGAGAACATCCCAAGACATTGGGMAAGTCACTATTTACCCCCACAGGGA 482
Qy GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
Db GAATTCGGTGTAGACAAAGCAGGATCCCCACTTTGTGAATTGCCTTATGTATAAAATG 542
Qy SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg 740
Db TCATACTACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCCAGGTTTGGACCGA 502
Qy ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe 760
Db ACACGTAAATGCTGAGATTGGAAATAAGGACATTAAATTCAAACATTTGGAAGAAGCCTTT 562
Qy ThrSerGluHisTyrLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
Db ACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCCACTGATTAACAGGGAGACA 722
Qy LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
Db TTAGATCACAACACCTCGAGTCACCAACATTTCCCAAAACAGAAATATTTGTCAAAGAAG 782
Qy ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
Db ACTACCAAAAGGAGCGGTGGCTACATTAAATAATAGCTGGTTTTTAAGAAAGGCAAGAA 842
Qy IleSerLysLysThrVal 826
Db ATATCTAAGAAGACTGTT 860
```

Search completed: December 14, 2004, 23:34:06
Job time : 988.69 secs

QY 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
 DB 240 TGGCTAAGGCAAAATACAGATGACATCAGAGTAAATGCTCTGGTGGGATTAATGGCTAT 299
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 629
 DB 300 CAGATAGCTGGGAATCGCTAATAGAACTACGTTGGTGGATATAACACCTGGAAATACAGC 359
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649
 DB 360 CACATAGCACTGGTGGGAAAGCTATGCTCTAATGAACAGCAGCTATAAATCATG 419
 QY 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
 DB 420 AGGACTCTAGATGATATGCTTTTGGTATTTTGGAGGGGTTTATGGCTATTCGGT 479
 QY 670 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyLysIlePheProLysAsp 689
 DB 480 GATGATATCAACAATTTCTCGGATGGTTAGGATAGCTGAAGGAGACATCCCAAGAC 539
 QY 690 IleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGlySer 709
 DB 540 ATTGGGAAAGTGACTATTTTACCCACAGGGAGAAATCCGTTAGACAAAGCAGGATCC 599
 QY 710 ProThrLeuAsnGlnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
 DB 600 CCTACTTTGTAATTCCTTATGATATAAAATGTCATCTACAGATTTGGGAAATGCG 659
 QY 730 LeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
 DB 660 CTGGATTTTCGTACACCCCGAGTTTGACCGACACAGCTAATGCTGAGATGGAAATAAG 719
 QY 750 AspIleLysPheLysIleGluGluAlaPheThrSerGluHisIleTrpLeuValArgIle 769
 DB 720 GACATTAATTCAAACATTTGGAAGAAGCCTTTACATCAGAACACCTGGCTTGTAGGATA 779
 QY 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysPheProArgValThrAsn 789
 DB 780 TATAAGTAAAGCACTGATACAGGGAGACATTAGATCACAACCTCGAGTCACCAAC 839
 QY 790 IlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIle 809
 DB 840 ATTTTCCCAAAACAGAAATTTGTCACAAAGAGACTACCAAAAGGAAGCGTGGCTACATT 899
 QY 810 LysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThrVal 826
 DB 900 AAAAATAAGCTGTTTATTAAGAAGGCAAGAAATATCTAAGAAGCTGTT 950
 RESULT 15
 AAH18021
 ID AAH18021 standard; cDNA; 2284 BP.
 XX
 AC AAH18021;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA sequence SEQ ID NO:17837.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX
 PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 EF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 PT
 PS Claim 8; SEQ ID NO 17837; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 2284 BP; 726 A; 388 C; 436 G; 734 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8.12e-284 Length: 2284
 Score: 286.00 Matches: 286
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 34.62% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-2 (1-826) x AAH18021 (1-2284)
 QY 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 DB 3 ATGTTGATGCTGATGCTATTGATGATGTTTGTCTCCATGTTACCTGGGTCACAGCAAT 62
 QY 561 AlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
 DB 63 GCCTACTCTAGTCCAAAGTAGTCCCTGCGCTCATACATCATGATGCGACCAAGATATC 122
 QY 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 DB 123 TTGATCATTTTGAAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACGA 182
 QY 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 DB 183 GTAATGCTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGCTTAATAGAACTACGTTG 242
 QY 621 ValAspAsnAenThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 DB 243 GTGGATAATAACACCTGGAAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTTCT 302
 QY 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660

```

Db 361 CACATAGCAGCTGGTGGGAAAGCTATGCTTCTTAATGAAACAGCAGCCTATAAATCATG 420
Qy 650 ArgThrLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 669
Db 421 AGGACTCTAGATGATAGATATGTTTGGTTATTTTGGAGGGGTATGCTCTATTCGGT 480
Qy 670 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 689
Db 481 GATGATATCAACAAATTTCTCGATGCTAGGATAGCTGAAGGAGAACATCCCAAGAC 540
Qy 690 IleArgGluSerAspTyrPheThrProGlyGlyGluPheArgValAspLysAlaGlySer 709
Db 541 ATTGGGGAAGTGACTATTTTACCCACAGGAGAAATTCGCTGTAGCAAAAGCAGATCC 600
Qy 710 ProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGln 729
Db 601 CCTACTTTGTTGATTCGCTTATGTATATAAATGTCATACATACAGATTTGGAGAAATGCAG 660
Qy 730 LeuAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLys 749
Db 661 CTGGATTTTCGTACACCCCGGCTTTTGCACGACACGTAATGCTGAGATTGGAATAAG 720
Qy 750 AspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIle 769
Db 721 GACATTAATCAACATTTTGGAGAGCTTTTACATCAGAACACTGGCTTGTAGGATA 780
Qy 770 TyrLysValLysAlaProAspAsnArgGluThrLeuAspHisLysProArgValThrAsn 789
Db 781 TATAAGTAAGACCCCTGTATACAGGAGACATTAGATCACAAACCTCGAGTCACCAAC 840
Qy 790 IlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIle 809
Db 841 ATTTCCCAAAACAGAGATTTTGTCAAGAGACATACCAAAAGGAGCGTGCTACATT 900
Qy 810 LysAsnLysLeuValPheLysLysGlyLysLysIleSerLysLysThrVal 826
Db 901 AAAAATAAGCTGTTTAAAGAAAGGCAAGAAATATCTAAGAAAGACTGTT 951

RESULT 14
AAD08289
ID AAD08289 standard; cDNA; 2537 BP.
XX
XX AAD08289;
XX
XX 08-AUG-2001 (first entry)
XX
XX Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 17.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
XX psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
XX inflammation; neurological disorder; Alzheimer's disease; food additive;
XX angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
XX pregnancy-related disorder; endocrine disorder; infection; wound healing;
XX cell culture; chemotaxis; vulnerability; binding partner identification;
XX gene therapy; ss.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 93..953
XX
XX
XX sig_peptide /tag= a
XX mat_peptide /product= "Human secreted protein precursor"
XX 93..155
XX 156..950
XX /tag= b
XX /tag= c
XX /product= "Mature human secreted protein"
XX
XX WO200136440-A1.

```

```

XX PD 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US031282.
XX
XX 19-NOV-1999; 99US-0166414P.
XX
XX 21-JUL-2000; 2000US-0219665P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Birse CE, Moore PA;
XX WFI; 2001-343795/36.
XX P-PSDB; AAE03824.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; Page 443; 553pp; English.
XX
XX AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
XX protein genes, and AAE03818-AAE03870 represent the proteins they encode.
XX AAE03871-AAE03896 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 23 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infections. The
XX proteins can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
XX present sequence represents a human secreted protein-encoding cDNA of the
XX invention
XX
XX Sequence 2537 BP; 816 A; 434 C; 487 G; 800 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. NO.: 4.2e-295 Length: 2537
XX Score: 297.00 Matches: 297
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 35.96% Indels: 0
XX DS: 4 Gaps: 0
XX
XX US-10-028-384-2 (1-826) x AAD08289 (1-2537)
XX
XX Qy 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetMet 549
XX Db 60 GGGTAGGCCCTAATATAAAGACATGTGCACCATGTTGATGCTATGATGATG 119
XX
XX Qy 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
XX Db 120 TTTGCTGTCCACTGTACCTGGGTCAACAGCATGCTACTCTAGTCCAGTGTAGCTCTG 179
XX
XX Qy 570 AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
XX Db 180 GCCTCATACATCATGATGGCACCAGGAATATCTTAGATGATTTTAGAGAGCTACTTTT 239

```

604 TrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
 Db TGGTGGGATATGCTATCAGATAGCTGGATGGCTAATAGAACTAGCTTGGTGGTAAAT 360
 624 AsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
 Db AACACCTGGAAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTCTTCTAATGAAACA 420
 644 AlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGlyGly 663
 Db CGAGCCTAATAATCANTAGAGACTCTAGATGTAGATTAATGTTTGGTTATTTTGGAGGG 480
 664 ValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGlu 683
 Db GTTATTGGCTATTCTGGTGTATGATATCAACAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 684 GlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArg 703
 Db GGAGAACATCCCAAGACATTCGGGAAGTGCATTTTACCCACAGGAGAAATTCCT 600
 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyr 723
 Db GTAGACAAAGCAGGATCCCTACTTCTGTAATTCCTTATGTATATAAATGTCACTAC 660
 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
 Db AGATTGGAGAAATCGAGCTGGATTTTCGTACACCCAGGTTTTCACCGAACACCTAT 720
 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPheThrSerGlu 763
 Db GCTGAGATTGGAATAAGACATTAATTAATTAATTCGAAGAGCCCTTACATCAGAA 780
 764 HisTrpLeuValAlaGlyTyrLysValLysAlaProAspAsnArgGluThrLeuAspHis 783
 Db CACTGGCTGTGTAGATATATAAGTAAAGCAACCTGATACAGGAGACATTAGATCAC 840
 784 LysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLys 803
 Db AAACCTCGAGTCACCAACATTTCCCAAAACAGAGATATTGTCAAAAGAGACTACCAAA 900
 804 ArgLysArgGlyTyrLysLysAsnLysLeuValPheLysLysGlyLysLysLysLys 823
 Db AGGAAGCGTGGCTACATTAATAATAGCTGGTTTTTAAGAAAGCAAGAAATATCTAAG 960
 824 LysThrVal 826
 Db AAGACTGTT 969

RESULT 13
 AAF93772 standard; cDNA; 2510 BP.
 XX
 AC AAF93772;
 DT 23-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding a membrane or secretory protein clone PSEC0070.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114090.
 XX
 PR 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 XX

PA (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPI; 2001-093989/11.
 DR P-PSDB; AAB88345.
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 XX gene therapy or as candidate target molecules in drug development.
 XX Claim 1; SEQ ID NO 57; 609pp + Sequence Listing; English.
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes
 XX

Sequence 2510 BP; 802 A; 412 C; 489 G; 807 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,06e-315 Length: 2510
 Score: 317.00 Matches: 317
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 38.38% Indels: 0
 DB: 5 Gaps: 0

US-10-028-384-2 (1-826) x AAF93772 (1-2510)

QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnLysThrGluGlu 529
 Db 1 TTGTATGATAGGACAGTAAAGTGGAGAAACATGCACTGAACAGAAAAAACTGAAGAG 60
 QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMet 549
 Db -61 GGATTAGGCCCTAATAATAAAGACATTGTCCACCATGTTGATGCTGATGCTATGATGATG 120
 QY 550 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeu 569
 Db 121 TTGTGCTGCACGTACCTGGGTCAACAGCAATGCTACTCTAGTCCAAAGTGTAGCTCTG 180
 QY 570 AlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPhe 589
 Db 181 GCCTCATCAATCATCATGCGCAGGAGATATCTTAGATGATTTTAGAGAGAGCTTACTT 240
 QY 590 TrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyr 609
 Db 241 TGGCTAGGCAAAATACAGATGAACATGACAGGATATGCTTGTGGATTAATGGCTAT 300
 QY 610 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 629
 Db 301 CAGATAGCTGGAATGGCTAATAGAACTACCTGGTGGATATAACACCTGGAATAACAGC 360
 QY 630 HisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMet 649

Db 121 CTGATGCTATTGATGATGTTTGCTGCACCTGTACCTGGGTCAACAGCAATGCTACTCT 180
Qy 564 SerProSerValValLeuAlaSerTyzAsnHisAspGlyThrArgAsnHisLeuAspAsp 583
Db 181 AGTCCAAGTGAGTCTGGGCTCATACATCATGATGCACAGGATATCTTAGATGAT 240
Qy 584 PheArgGluAlaTyPheThrLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603
Db 241 TTTAGAGAGCTTACTTTTGCTTAAGGCAAAATACAGATGAACATGCACGAGTAATGTCT 300
Qy 604 TrpTrpAspTyGlyTyGlnHisAlaGlyMetAlaAsnArgThrThrLeuValAspAsn 623
Db 301 TGGTGGGATATGCTATCATGATAGCTGGATGGCTATATGAGACTAGCTTGGGGATAT 360
Qy 624 AsnThrTrpAsnAsnSerHisLeAlaLeuValGlyLysAlaMetSerSerAsnGluThr 643
Db 361 AACACCTGGGAATACAGCCCATAGCACTGGTGGGAAAGCTATGCTCTTAATGAACA 420
Qy 644 AlaAlaTyLysLysLeuMetArgThrLeuAspValAspTyPheValLeuValLlePheGly 563
Db 421 GCAGCTTATTAATCATGAGGACTCTAGATGTAGATTAATGTTTGGTTATTTTGGAGG 480
Qy 564 ValIleGlyTySerGlyAspAspIleAsnLysPheLeuTrpMetValArgLleAlaGlu 683
Db 481 GTTATTGGCTATTCTGGTGTATGATATCAACAAATTTCTCTGGATGTTAGGATAGTGA 540
Qy 684 GlyGluHisProLysAspIleArgGlySerAspTyPheThrProGlnGlyLlePheArg 703
Db 541 GGAGACATCCCAAGACATCCGGAAAGTGACTATTTTACCCACAGGAGAAATTCCT 600
Qy 704 ValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyLysMetSerTyTy 723
Db 601 GTAGACAAAGCAGGATCCCTACTTTGTTGATTTGCTTATGTATATAAATGTCACTAC 660
Qy 724 ArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsn 743
Db 661 AGATTGGAGAAAGTGGCTGGATTTTGGTACACCCCGAGTTTGGACCGAACACGTAT 720
Qy 744 AlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGlu 763
Db 721 GCTGAGATTGGAATAAAGACATTAATTCACATTTGGAAGAGCCCTTTACATCAGNA 780
Qy 764 HisTrpLeuValArgIleTyLysValLysAlaProAspAsnArgGluThrLeuAspHis 783
Db 781 CACTGCTGTTAGGATATATAAAGTAAAGCACCTGTATAACAGGAGACATTAGATCAC 840
Qy 784 LysProArgValThrAsnIlePheProLysGlnLysTyLeuSerLysLysThrThrLys 803
Db 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 900
Qy 804 ArgLysArgGlyTyTrileLysAsnLysLeuValPheLysLysGlyLysLysSerLys 823
Db 901 AGGAAGCGTGGCTACATTAAATAAATAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAG 960
Qy 824 LysThrVal 826
Db 961 AGACCTGTT 969

RESULT 12

AAF98463
ID AAF98463 standard; cDNA; 2546 BP.
AC AAF98463;
XX AAF98463;
DT 07-JUN-2001 (first entry)
XX Human cDNA clone CT585_1 sequence SEQ ID 150.
DE Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW haematopoiesis.

XX Homo sapiens.

XX WO2001.19988-A1.

XX 22-MAR-2001.

XX 14-SEP-2000; 2000WO-US025135.

XX 17-SEP-1999; 99US-00398829.

XX (GENY) GENETICS INST INC.

XX Jacobs K, McCoey JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;

XX WPI: 2001-244801/25.

XX P-PSDB; AAB90727.

XX Isolated nucleic acids encoding polypeptides, useful for modulating e.g.
PT cytokine and cell proliferation/differentiation activity, the immune
PT system and hematopoiesis regulating activity.

XX Disclosure; Page 476-477; 557pp; English.

XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokinetic activity; haemostatic and thrombolytic activity;
CC receptor/ligand activity; anti-inflammatory activity; haematopoiesis
CC activity; cadherin/tumour suppressor activity; and/or tumour inhibition
CC activity. Included in the invention are probes represented in AAF98490 -
CC AAF98572 which are specific for the cDNA clones encoding the secreted
CC proteins

XX SQ Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2546
Score: 323.00 Matches: 323
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.10% Indels: 0
DB: 5 Gaps: 0

US-10-028-384-2 (1-826) x AAF98463 (1-2546)

Qy 504 LysArgAsnGlnGlyAsnLeuTyAspLysAlaGlyLysValArgLysHisAlaThrGlu 523

Db 1 AAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAGGAGGAAATGCACTGAA 60

Qy 524 GlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543

Db 61 CAGGAAAAAATCTGAAGGAGGATTTAGCCCTATATATAAAGCAATTTCCACATGTTGATG 120

Qy 544 LeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTySer 563

Db 121 CTGATGCTATTGATGATGTTTGTCTGTCCACTGTACCTGGGTCAACAGCAATGCCTACTCT 180

Qy 564 SerProSerValValLeuAlaSerTyAsnHisAspGlyThrArgAsnIleLeuAspAsp 583

Db 181 AGTCCAAAGTGTAGTCTCTGGCTCATCAATCATGATGATGATGATGATGATGATGAT 240

Qy 584 PheArgGluAlaTyPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSer 603

Db 241 TTTAGAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCACAGATGATGCT 300

```

ID 2 TTGCGCGTCATCCGCTTGAAGACATCATCCAGGTTCCGACCGTGGTTAACTATAGA 61
XX
AC
XX
XX 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAenTrpPheAspGluArg 129
XX
XX 62 TCACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATGGTTGATGAAAGA 121
XX
XX 130 AlaTrpTyrProLeuGlyArgGlyValGlyGlyThrValTyrProGlyLeuMetIleThr 149
XX
XX 122 GCATGGTATCCACTAGGAAGAAATAGTGGTACTGTTTACCCAGGTTGATGATACC 181
XX
XX 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAenIleThrValHisIleArgAspVal 169
XX
XX 182 GTGGCCATTATTCATGGATTTTAAATACATTGAACATAACTGTTTCACATAAGACGTA 241
XX
XX 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
XX
XX 242 TGTGTGTTCCCTGCAACATTTTATAGCGCCCTTACATCTATATCTACTTCTCTGTAC 301
XX
XX 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
XX
XX 302 AGAAGACATTTGGACCAAGACAGGACTTTTAGCTGCTGTTTATTGCTATTGTACCA 361
XX
XX 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
XX
XX 362 GGCTACATATCTCGGTGAGTGTGATCTTTGATAATGAAGGCAITGCTATTTTTCGA 421
XX
XX 230 LeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheTrpThr 249
XX
XX 422 CTTCACTTACATCTATTATGTTGGTAAATCTGTAAACCTGGTCTGCTTTTGGACA 481
XX
XX 250 MetCysCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 269
XX
XX 482 ATGTGCTGCTTATCTCTATTTCTATATGTTCTCTGCTTGGGTGGTATTATTTATC 541
XX
XX 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArg 289
XX
XX 542 ATCAATCTTATTCATCTGATGATGATTGTTGTTGTTACTGATGAGATACAGCAAGA 601
XX
XX 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeuSerMetGlnIlePro 309
XX
XX 602 GTTACATAGCATATAGCACTTCTACATTTGGGTTTAAATATTATCATGCAATACCT 661
XX
XX 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
XX
XX 662 TTTGTGGGATTCAGGCAATCAGAACAGTGAAACATGCGAGCTGCGAGTGTCTTTGCA 721
XX
XX 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
XX
XX 722 TTGCTGCAAGCTTAAGCTTTCTTGCAGTATCTGAGAGACCGATTAAACAAACAGAGTTC 781
XX
XX 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
XX
XX 782 CAGACCTTTCTTTTGGGTGATCATCACTAGCTGCAGGTGCTGTCTTCTTAGTGTATC 841
XX
XX 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
XX
XX 842 TATTGTGCTTATACAGGTTCATATGCACCATGAGTGGCAGGTTTATTCATTGTTGGAT 901
XX
XX 390 ThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnProThr 409
XX
XX 902 ACTGGGTATGCAAAATATACATTCCAATATTGTCATCAGTGTCTGAGCATCAACCTAGC 961
XX
XX 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
XX
XX 962 ACTTGGGTGCTTTCTTCTTTGATCTACATATTCTTGTATGTACCTTCCAGCGACCTT 1021
XX
XX 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
XX
XX 1022 TGGTTCGTCATCAAAATATATCAACGATGAAAGA 1054
XX

```

RESULT 11
AAV44866

```

ID AAV44866 standard; cDNA; cDNA; 2546 BP.
XX
AC AAV44866;
XX
XX 21-OCT-1998 (first entry)
XX
XX Clone CT585_1 coding sequence.
XX
XX Secreted protein; nutritional source; cell proliferation activity;
XX cell differentiation activity; immune stimulant; tissue growth activator;
XX haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
XX tumour inhibitor; clone CT585_1; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 112..972
XX /*tag= a
XX
XX WO9825962-A2.
XX
XX 18-JUN-1998.
XX
XX 12-DEC-1997; 97WO-US023224.
XX
XX 13-DEC-1996; 96US-00766263.
XX
XX 11-DEC-1997; 97US-00989232.
XX
XX (GEMY ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
XX Spaulding V, Agostino MG,
XX WPI; 1998-362424/31.
XX DR P-PSDB; AAW69247.
XX
XX New isolated polynucleotides - obtained from human adult testis, human
XX adult ovary, human adult brain and human adult heart cDNA libraries.
XX
XX Claim 35; Page 79-81; 108pp; English.
XX
XX This sequence represents a polynucleotide of the invention, and encodes a
XX secreted protein. It was isolated from a human adult brain cDNA library,
XX and is designated clone CT585_1. The DNA sequences and encoded
XX polypeptides can be used as nutritional sources or supplements, or may
XX exhibit e.g. cytokine and cell proliferation/differentiation activity,
XX immune stimulating or suppressing activity, haematopoiesis regulating
XX activity, receptor/ligand activity, anti-inflammatory activity,
XX activin/inhibin activity, chemostatic/chemokinetic activity,
XX cadherin/tumour invasion suppressor activity, tissue growth activity,
XX tumour inhibition activity or other activities
XX
XX Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;
XX

```

Alignment Scores:

| Pred. No.: | Length: |
|------------------------|---------|
| Score: | 323.00 |
| Percent Similarity: | 100.00% |
| Best Local Similarity: | 100.00% |
| Query Match: | 39.10% |
| DB: | 2 |
| | Gaps: |
| | 0 |

US-10-028-384-2 (1-826) x AAV44866 (1-2546)

```

QY 504 LysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGlu 523
DB 1 AAAGAACACCAAGGAATTTGATGATAAGGCAGGTAAAGTGAGAAACATGCAACTGAA 60
QY 524 GlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMet 543
DB 61 CAGGAAAAAACTGAAGAGGGGATTAGGCCCTTAATAAAAAAGCATGTGCACCATGTTG 120
QY 544 LeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSer 563

```

QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 362 GGCTACATATCTCGGTAGTGGATCCCTTTCATATGAGGCAATGCTATTTTGA 421
 QY 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleThrGlySerValPheTrpThr 249
 Db 422 CTTTCAGTTTCACATATTTATGGGTAAATCTGTAAACCTGGGTGAGTTTTTTGGACA 481
 QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
 Db 482 ATGTGCTGCTGCTTATCTATTTCTATATGCTCTCTGCTGGGTGCTTATGATTTATC 541
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerIleArg 289
 Db 542 ATCAATCTTATTCCTACTGATGATTTGTGTGTGTCTGATGACAGATACACAAAGA 601
 QY 290 ValTyrIleAlaTyrSerPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 Db 602 GTCTACATAGCATATGACATTTCTACATTTGCTGGGTAAATATTTATCAATGCAGATACCT 661
 QY 310 PheValGlyPheClnProIleArgThrSerGluHisMetAlaIleAlaGlyValPheAla 329
 Db 662 TTGTGGGATTCACCCCAATCAGAACAGTGAACATGCGAGCTGCGAGTGTCTTGA 721
 QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnGluPhe 349
 Db 722 TTGCTGCAAGCTTATGCTTTCTTCTGAGTATCTGAGAGACCGATTTAAACAAACAGAGTTC 781
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaIleAlaGlyValPheLeuSerValIle 369
 Db 782 CAGACCCCTTTCTTTGGGTGATCCTAGCTCAGGTGCTGTGTTCTTATGATGATC 841
 QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAsp 389
 Db 842 TATTTGACTTATACAGGTATACATTCACCATGGAGTGGCAGGTTTATTCATTTGGGAT 901
 QY 390 ThrGlyTyrAlaValIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 Db 902 ACTGGTATGCAAAATACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 961
 QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 Db 962 ACTTGGGTGCTTTCTTCTTGTATCTACATATTCCTGATGATGATGATGATGATGATG 1021
 QY 430 TrpPheCysIleLeuAsnIleAsnAspGluArg 440
 Db 1022 TGGTTCTGATCAAAATATCAACATGAAGA 1054
 RESULT 10
 ID ADN60663
 AC ADN60663 standard, cDNA; 1543 BP.
 XX ADN60663;
 XX
 DT 01-JUL-2004 (first entry)
 DE Human secreted polynucleotide #123.
 XX Human; secreted polynucleotide; gene; ss; autoimmune disease;
 KW rheumatoid arthritis; hyperproliferative disorder; neoplasm; breast;
 KW liver; cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; nervous system disorder; Alzheimer's disease;
 KW bacterial infection; viral infection; fungal infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin aging; sunburn; organ transplantation; tissue regeneration;
 KW chemotaxis; food additive; food preservative; fat content; vitamin;
 KW mineral.
 XX
 OS Homo sapiens.
 XX
 PN US2004038277-A1.
 XX

PD 26-FEB-2004.
 XX 18-JUL-2003; 2003US-00621401.
 XX 07-NOV-1997; 97US-0064900P.
 PR 07-NOV-1997; 97US-0064908P.
 PR 07-NOV-1997; 97US-0064911P.
 PR 07-NOV-1997; 97US-0064912P.
 PR 07-NOV-1997; 97US-0064983P.
 PR 07-NOV-1997; 97US-0064984P.
 PR 07-NOV-1997; 97US-0064985P.
 PR 07-NOV-1997; 97US-0064987P.
 PR 07-NOV-1997; 97US-0064988P.
 PR 07-NOV-1997; 97US-0066089P.
 PR 17-NOV-1997; 97US-0066090P.
 PR 17-NOV-1997; 97US-0066094P.
 PR 17-NOV-1997; 97US-0066095P.
 PR 17-NOV-1997; 97US-0066100P.
 PR 04-NOV-1998; 98WO-US023435.
 PR 05-MAY-1999; 99US-00305736.
 PR 13-OCT-2000; 2000US-0239893P.
 PR 28-MAR-2001; 2001US-00818683.
 PR 12-OCT-2001; 2001US-00974879.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Feng P, Ruben SM, Ebner R, Olsen H, Ni J, Wei Y;
 PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
 PI Endress GA, Carter KC, Birse CE;
 XX WPI; 2004-203215/19.
 DR P-PSDB; ADN60791.
 XX
 PT New nucleic acid molecule encoding one of 125 human secreted proteins for
 diagnosing, preventing, treating or ameliorating medical conditions and
 used as food additives or preservatives.
 XX
 PS Claim 1; SEQ ID NO 133; 531bp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules encoding human
 secreted proteins. The polynucleotides and polypeptides are used in
 diagnosing a pathological condition or susceptibility to a pathological
 condition. Antibodies to the polypeptides can also be used in alleviating
 symptoms associated with the disorders and in diagnostic immunoassays
 (e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA)).
 CC Disorders which are diagnosed or treated include autoimmune diseases
 (e.g. rheumatoid arthritis), hyperproliferative disorders (e.g. neoplasms
 of the breast or liver), cardiovascular disorders (e.g. cardiac arrest),
 CC cerebrovascular disorders (e.g. cerebral ischaemia), nervous system
 disorders (e.g. Alzheimer's disease), infections caused by bacteria,
 CC viruses and fungi and ocular disorders (e.g. corneal infection). The
 CC polypeptides can also be used to aid wound healing and epithelial cell
 proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues and in chemotaxis. The polypeptides can also be
 CC used as food additives or preservatives to increase or decrease storage
 CC capabilities and fat content, as vitamins, as minerals and as other
 CC nutritional components. This sequence represents a human secreted
 CC polynucleotide of the invention.
 XX
 SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 1543
 Score: 351.00 Matches: 351
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 42.49% Indels: 0
 DB: 12 Gaps: 0
 US-10-028-384-2 (1-826) x ADN60663 (1-1543)

QY 90 PheAlaValIleArgPheGluSerIleHisGluPheAspProTrpPheAsnTyrArg 109

(SHI Y.)
(JANAT F.)
(ENDS) (ADDRESS G A.)
(CARTER K C.)

Feng P, Ruben SM, Rosen CA, Ebner R, Olsen HS, Ni J, Wei Y;
Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
Endress GA, Carter KC;

WPI: 2004-051575/05.
P-PSDB; ADG78500.

New secreted nucleic acid for diagnosing, preventing or treating diseases
associated with aberrant expression or activity of the polypeptide it
encodes: e.g. cancer, human immunodeficiency virus, Parkinson's disease,
or diabetes.

Claim 1: SEQ ID NO 133; 377pp; English.

The invention also relates to human secreted proteins and the nucleic
acids encoding them. The proteins and nucleic acids are useful in
diagnosing, preventing, prognosing or treating diseases or disorders
associated with aberrant expression and/or activity of the secreted/
proteins, such as immune disorders, haematopoietic disorders,
hyperproliferative disorders, infectious diseases or inflammatory
disorders. In particular, the diseases or disorders are HIV, anaemia,
thrombocytopenia, bleeding, stroke, myocardial infarction, Addison's
disease, rheumatoid arthritis, dermatitis, Grave's disease, multiple
sclerosis, glomerulonephritis, diabetes, graft-versus-host disease,
inflammatory bowel disease, cancer, bacterial infections, viral
infections, parasitic infections, osteoporosis, wounds, atherosclerosis,
Alzheimer's disease, Parkinson's disease or ulcers. The sequences may
also be used as food additives or preservatives, or for modulating
mammalian mental or physical characteristics. The nucleic acids are also
used in chromosome mapping, in forensic biology or as molecular weight
markers. This sequence represents cDNA encoding a human secreted protein
of the invention.

Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Alignment Scores:

| Pred. No.: | Score: | Length: | 1543 |
|------------------------|---------|---------------|------|
| Percent Similarity: | 351.00 | Matches: | 351 |
| Best Local Similarity: | 100.00% | Conservative: | 0 |
| Query Match: | 100.00% | Mismatches: | 0 |
| | 42.49% | Indels: | 0 |
| DB: | 12 | Gaps: | 0 |

US-10-028-384-2 (1-826) x ADG78372 (1-1543)

| | | | |
|----|-----|--|-----|
| Qy | 90 | PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyArg | 109 |
| Db | 2 | TTTCGGCGTCATCCGCTTCGAAAGCATCATCCAGGTCACCGCGTGTAACTATAGA | 61 |
| Qy | 110 | SerThrHisHisLeuAlaSerHisGlyPheTyArgPheLeuAsnTrpPheAspGluArg | 129 |
| Db | 62 | TCAACACATCATCTTGCACTCATGGGTCTATCAATTTTAAATGGTTGATGAAGA | 121 |
| Qy | 130 | AlaTrpTrpProLeuGlyArgIleValGlyThrValTyProGlyLeuMetIleThr | 149 |
| Db | 122 | GCATGGGTATCCACATAGGAAGAATAAGTAGGTGGTACTGTTTACCACGGTGTATGATPACC | 181 |
| Qy | 150 | AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal | 169 |
| Db | 182 | GCTGGCCTTATTCATTGGATTTTAAATACATTGAACATACTGTTCCATACAGACGTA | 241 |
| Qy | 170 | CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr | 189 |
| Db | 242 | TGTGTGTTCCCTTGACCAACATTTTAGCGCGCTTACATCTATACTTCTCTCTTACA | 301 |
| Qy | 190 | ArgGluLeuTrpAsnGlnGlyValGlyLeuAlaAlaCysPheIleAlaIleValPro | 209 |
| Db | 302 | ACAGAACTTTGGAACTCAGGACGACGACTTTAGCTGCTTTGTTTATGCTATTGTACCA | 361 |

RESULT 7

AAx85055
ID AAX85055 standard; DNA; 1543 BP.

XX AC AAX85055;

XX DT 30-JUL-1999 (first entry)

XX DE Human secreted protein gene No. 123.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO924836-A1.

XX PD 20-MAY-1999.

XX PF 04-NOV-1998; 98WO-US023435.

XX PR 07-NOV-1997; 97US-0064900P.

XX PR 07-NOV-1997; 97US-0064908P.

XX PR 07-NOV-1997; 97US-0064911P.

XX PR 07-NOV-1997; 97US-0064912P.

XX PR 07-NOV-1997; 97US-0064983P.

XX PR 07-NOV-1997; 97US-0064984P.

XX PR 07-NOV-1997; 97US-0064985P.

XX PR 07-NOV-1997; 97US-0064987P.

XX PR 07-NOV-1997; 97US-0064988P.

XX PR 17-NOV-1997; 97US-0066089P.

XX PR 17-NOV-1997; 97US-0066090P.

XX PR 17-NOV-1997; 97US-0066094P.

XX PR 17-NOV-1997; 97US-0066095P.

XX PR 17-NOV-1997; 97US-0066100P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Feng P, Carter KC, Endress GA, Rosen CA, Ruben SM, Janat F;

XX PI Ni J, Wei Y, Moore PA, Soppet DR, Kyaw H, Lafleur DW, Olsen HS;

XX PI Shi Y, Ebner R;

XX DR WPI; 1999-337740/28.

XX DR P-PSDB; AAY27689, AAY27918, AAY27919, AAY27920, AAY27921, AAY27922,

XX DR AAY27923.

XX PT New human secreted proteins and coding sequences useful for treating

XX PT disorders of the immune system and hyperproliferative disorders.

XX PS Claim 1; Page 347; 507pp; English.

XX CC This sequence represents a nucleic acid molecule which encodes a secreted

XX CC human protein. The gene number is given in the descriptor line. The gene

XX CC can be used to generate fusion proteins by linking to the gene to a human

XX CC immunoglobulin Fc portion (e.g. AAX84924) for increasing the stability of

| Alignment Scores: | | Pred. No.: | 0 | Length: | 1543 |
|---|---------|---|-----|---------|------|
| Score: | 351.00 | Matches: | 351 | | |
| Percent Similarity: | 100.00% | Conservative: | 0 | | |
| Best Local Similarity: | 100.00% | Mismatches: | 0 | | |
| Query Match: | 42.49% | Indels: | 0 | | |
| DS: | 2 | Gaps: | 0 | | |
| US-10-028-384-2 (1-826) x AAX85055 (1-1543) | | | | | |
| Qy | 90 | PhaAlaValIleArgPheGluSerIleHISGluPheAspProTrpPheAsnTrpArg | 109 | | |
| Db | 2 | TTGGCGGTCACTCCGCTTCGAAGACATCCACAGATTCGACCCGTGGTTTAACTATGAGA | 61 | | |
| Qy | 110 | SerThrHisHisLeuAlaSerHisGlyPheTrpGluPheLeuAsnTrpPheAspGluArg | 129 | | |
| Db | 62 | TCAACACATCATCTTCATCTCATGGTTCTATGAATTTTAAATGGTTTGAAGAGA | 121 | | |
| Qy | 130 | AlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTrpProGlyLeuMetIleThr | 149 | | |
| Db | 122 | GCATGGTATCCACTAGGAAGAATAGTAGGTGTGTATCTTTACCCAGGTTGATGATAACC | 181 | | |
| Qy | 150 | AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal | 169 | | |
| Db | 182 | GCTGGCCTTATTCATTGGATTTTAAATACATTTGAACATACTGTTTCACATTAAGAGACGA | 241 | | |
| Qy | 170 | CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr | 189 | | |
| Db | 242 | TGTGTGTCTCTTGCACCACTTTTAGCGGCCCTTACATCTATATCTACTTCTCTCTTACA | 301 | | |
| Qy | 190 | ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValPro | 209 | | |
| Db | 302 | AGAGAACTTTGGAAACCAAGGACGAGACTTTTAGCTGCTTTTATTGCTATTGTACCA | 361 | | |
| Qy | 210 | GlyTrpIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla | 229 | | |
| Db | 362 | GGCTACATATCTCGTCAGTAGCTGATCTTTGATATGATGAAGGCATTCCTATTTTTGA | 421 | | |
| Qy | 230 | LeuGlnPheThrTrpTrpLeuTrpValIleSerValIleThrGlySerValPheTrpThr | 249 | | |
| Db | 422 | CTTCAGTTTCACATACATATTTATGGGTAAATCTGTAAATACTGGGTGCTGTTTGGACA | 481 | | |
| Qy | 250 | MetCysCysCysLeuSerTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPheIle | 269 | | |
| Db | 482 | ATGTGCTGCTGCTTATCTCTATATATGGTCTCTGCTTGGGGTGGTATGATTTATC | 541 | | |
| Qy | 270 | IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTrpSerIleArg | 289 | | |
| Db | 542 | ATCAATCTTATTCACCTGCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG | 601 | | |
| Qy | 290 | ValTrpIleAlaTrpSerThrPheTrpIleValGlyLeuIleLeuSerMetGlnIlePro | 309 | | |
| Db | 602 | GTCATACATAGCATATAGCATCTTCTACATGTTGGGTTTAAATATTATCATCATGATACCT | 661 | | |
| Qy | 310 | PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla | 329 | | |
| Db | 662 | TTTGTGGATTTCAGCAATCAGAACACTGAAACATGGCAGCTGCAGGTTGCTTTTGA | 721 | | |
| Qy | 330 | LeuLeuGlnAlaTrpAlaPheLeuGlnTrpLeuArgAspArgLeuThrIleGluPhe | 349 | | |
| Db | 722 | TTGCTGCAAGCTTATGCTTCTTTCAGATATCTGAGAGACCGATTAACAAAACAGAGATC | 781 | | |
| Qy | 350 | GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle | 369 | | |
| Db | 782 | CAGACCTTTCTTTTGGGTGTATCCTAGCTAGCTGAGGTGCTGTGCTTCTTAGTGTATC | 841 | | |
| Qy | 370 | TyrLeuThrTrpThrGlyTrpIleAlaProTrpSerGlyArgPheTrpSerLeuTrpAsp | 389 | | |
| Db | 842 | TATTTGACTTATACAGGTACATTTGACCATGGAGTGGCAGGTTTATTCATTGTTGGAT | 901 | | |
| Qy | 390 | ThrGlyTrpAlaIleHisIlePheIleAlaSerValSerGluHisGlnProThr | 409 | | |

QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1022 TGGTTCATCAAAATATCAACGATGAAGA 1054
 RESULT 6
 ABL89850
 ID ABL89850 standard; cDNA; 1209 BP.
 AC ABL89850;
 DT 24-MAY-2002 (first entry)
 XX Human polynucleotide SEQ ID NO 412.
 DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW anti-allergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; gene; ss.
 OS Homo sapiens.
 XX
 XX WO200190304-A2.
 XX
 XX 29-NOV-2001.
 XX
 XX 18-MAY-2001; 2001WO-US016450.
 XX
 XX 19-MAY-2000; 2000US-0205515P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Birse CE, Rosen CA;
 FI
 XX WPI; 2002-122018/16.
 DR
 DR P-PSDB; ABL89441.
 XX
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX
 XX Claim 4; SEQ ID NO 412; 2081pp + Sequence Listing; English.
 PS
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABL89040-ABL90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX
 XX Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 1209
 Score: 351.00 Matches: 351
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 42.49% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-2 (1-826) x ABL89850 (1-1209)

QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTTPheAsnTyrArg 109
 DB 2 TTGGCGGTTCATCGGTTTGGAAAGCATCATCCAGAGTTCGACCCGTGGTTTAACATATAGA 61
 QY 110 SerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTTPheAspGluArg 129
 DB 62 TCAACACATCATCTTGCATCTCATGGTTCATGAATTTTAAATGGTTGATGAAGA 121
 QY 130 AlaTTPTrpProLeuGlyArgIleValGlyGlyThrValTyrProGlyIleuMetIleThr 149
 DB 122 GCATGGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCCAGGGTTGATGATAACC 181
 QY 150 AlaGlyLeuIleHisTTPTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 DB 182 GCTGGCCTTATTCATGGATTAAATACATTGAACATACTGTTTACATTAAGAGCGTA 241
 QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 DB 242 TGTGTGTTCTTGCACCAACTTTTAGCGGCTTATCATCTATATCTACTTCTCTCTTACA 301
 QY 190 ArgGluLeuTrpAsnGlnGlyValAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 DB 302 AGAGAACTTTGGAACCAAGAGGAGGAGACTTTTAGTCTGCTTGTTTATTTGTTATGACCA 361
 QY 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 DB 362 GGTACATATCTCGGTCTAGTAGTCTTTCATTAATGAAGGCATTGCTTATTTTGCA 421
 QY 230 LeuGlnPheThrTyrTyrLeuTTPValLysSerValLysThrGlySerValPheThrThr 249
 DB 422 CTTCAGTTTCATACATCTATTATGGTAAATCTGTAAAACTGGGTGAGTTTGTGACA 481
 QY 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIle 269
 DB 482 ATGTGCTGCTCTTATCTATATGCTCTCTCTGGGTGGTGTATGATTTATC 541
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
 DB 542 ATCATCTTATCCACTGCAATGTAATTTGTGTTGTTACTGATGAGAGATACAGCAAGA 601
 QY 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIlePro 309
 DB 602 GTCTACATAGCATATAGCACTTTCTACATTTGGTGGTTTAAATATATCAATGCAGATACCT 661
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAla 329
 DB 662 TTTGTGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTGCGAGTGTCTTTGCA 721
 QY 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
 DB 722 TTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTACAAACACAGAGTTC 781
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 DB 782 CAGACCCCTTTCTTTTGGGTGTATCACTAGCTGCGAGTGTCTGTTTCTTGTAGTGTATC 841
 QY 370 TyrLeuThrTyrThrGlyTyrIleAlaProTTPSerGlyArgPheTyrSerLeuTTPAsp 389
 DB 842 TATTTGATCTATACAGGTACATTTGACCATGAGTGGAGTGGCAGGTTTATTCATTGTGGAT 901
 QY 390 ThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 DB 902 ACTGGGTATGCAAAATATACATTTCCAATTTATGTCATGAGTGTCTGAGCATCAACCTACG 961
 QY 410 ThrTTPValSerPhePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 DB 962 ACTTGGGTGTCTTCTTCTTGTATCTACATATTTCTTGTATGATACCTTCCACAGAGCCTT 1021
 QY 430 TrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1022 TGGTTCATCAAAATATCAACGATGAAGA 1054

Db 2263 TGCACCTCCCGAGCTTTGACCAACACGTAATGCTGAGATTGGAAATTAAGACATTAA 2322
 QY 752 sPhelyHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgileTyrlsVa 772
 Db 2323 ATTCAAGCATTTGGAGGAGCTTTTACATCAGAGCACTGGCTTGTGAGATATATAAGT 2382
 QY 772 llyshlaProAspAsnArgGluThrLeu 781
 Db 2383 GAAAGCACCTGACACACGGGAGACACTA 2410
 RESULT 5
 AAH33264
 ID AAH33264 standard; cDNA; 1209 BP.
 XX
 AC AAH33264;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:320.
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US026524.
 XX
 PR 29-SEP-1999; 98US-0157137P.
 PR 03-NOV-1999; 99US-0163280P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC..
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX WPI; 2001-235357/24.
 DR P-PSDB; AAG73833.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers.
 XX
 PS Claim 1; Page 2436; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
 CC proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene therapy
 CC and vaccine production. N and P may be used in the prevention, diagnosis
 CC and treatment of diseases associated with inappropriate P expression. For
 CC example, N and P may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of P by expressing inactive proteins or to
 CC supplement the patients own production of P. Additionally, N may be used
 CC to produce the colon cancer-associated ps, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the proteins. N and P
 CC can be used in the prevention, diagnosis and treatment of colorectal
 CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
 CC sequences used in the exemplification of the present invention. N.B.
 CC Pages 866 to 882 and page 7053 of the sequence listing were missing at
 CC time of publication, meaning no sequences are present for SEQ ID NO:1027
 CC to 1052, 7921 and 7922
 XX
 SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 1209
 Score: 351.00 Matches: 351
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 42.49% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-2 (1-826) x AAH33264 (1-1209)
 QY 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArg 109
 Db 2 TTGCGCGTTCATCCGCTTCGAAAGCATCATCCACGAGTTCGACCCGCTGGTTTAACTATAGA 61
 QY 110 SerThrHisLeuAlaSerHisGlyPheTyrlGluPheLeuAsnTrpPheAspGluArg 129
 Db 62 TCAACACATCATCTTCATCTCATGGTTCATGAATTTTAAATGGTTCATGAAAGA 121
 QY 130 AlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTyrlProGlyLeuMetIleThr 149
 Db 122 GCATGGTATCCACTAGAGAAATAGTAGGTGGTACTGTTTACCAGGGTTGATGATAACC 181
 QY 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 Db 182 GCTGGCGCTTATTCATTTGGATTTTAAATACATTTGAACATTAACCTGCATTAAGACACGTA 241
 QY 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 242 TGTGTGTTCCTTCACCAACTTTTAGCGGCCCTTACATCTATATCTACTTCTCTGCTTACA 301
 QY 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 Db 302 AGAGAACTTTGGAAACCAAGAGCAGGACCTTTTAGCTGCTGTTTATTTATTTGCTATTGTA 361
 QY 210 GlyTrpIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 362 GGCTACATATCTCGTCAGTAGCTGATCTCTTGATATGAAGGCAATGCTATTTTTGA 421
 QY 230 LeuGlnPheThrTyrlTyrlLeuTrpValIleValIleValIleValIleValIleVal 249
 Db 422 CTTTCAGTTTACATCTATTTATGGGTAAATCTGTAAAACTGGGTGAGTTTGTGGACA 481
 QY 250 MetCysCysLeuSerTyrlPheTyrlMetValSerAlaTrpGlyGlyTyrlValPheIle 269
 Db 482 ATGTGCTGCTGCTTATCTTCTATATGCTCTCTGCTGGGGTGGTATGATTTATTC 541
 QY 270 IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrlSerIleVal 289
 Db 542 ATCAATCTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
 QY 290 ValTyrlleAlaTyrlSerThrPheTyrlleValGlyLeuIleLeuSerMetGlnIlePro 309
 Db 602 GTCTACATAGCATATAGCACTTCTTACATTTGGGGTTTAAATATATCAATGCAGATACCT 661
 QY 310 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAla 329
 Db 662 TTTGTGGATTCCAGCAATCAGAACAGTGAACATGCGCAGCTGCGAGTGTCTTTGCA 721
 QY 330 LeuLeuGlnAlaTyrlAlaPheLeuGlnTyrlLeuArgAspArgLeuThrIleGlnIlePhe 349
 Db 722 TTCTGCAAGCTTATCTTCTTGTGAGTATCTGAGAGACCGGATTAACAAAACAGAGTTC 781
 QY 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIle 369
 Db 782 CAGACCTTTCTTTTGGGTGTACATCTAGCTGAGGTGCTGCTGCTTCTTAGTGTCTATC 841
 QY 370 TyrlLeuThrTyrlThrGlyTyrlleAlaProTrpSerGlyArgPheTyrlSerLeuTrpAsp 389
 Db 842 TATTTGACTTATACAGGTATACATTTGACCATGGAGTGGCAGGTTTTTATTCATTTGGGAT 901
 QY 390 ThrGlyTyrlAlaIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 Db 902 ACTGGGTATGCAAAAATACACATTTCAATTAATGATCAGTGTCTGAGCATCAACCTACG 961
 QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 Db 962 ACTTGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1021

Pred. No.: 0 Length: 2710
 Score: 365.00 Matches: 720
 Percent Similarity: 98.0% Conservative: 0
 Best Local Similarity: 98.0% Mismatches: 7
 Query Match: 44.1% Indels: 14
 DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x ADD94785 (1-2710)

| | | | |
|----|------|---|------|
| QY | 55 | ProAlaGlyLeuSerGlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPhe | 74 |
| Db | 230 | CCCGGGGGCTGTCGGGGGCTTGTGGAGCGGCGGGTGGAGTGGTGGCTCTCTC | 289 |
| QY | 75 | ThrIleLeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArg | 94 |
| Db | 290 | ACATCTCTCTCTGGCCGCTGGCCGCTTCAGCTCGCCCTCTCTGGCCGCTCATCCG | 349 |
| QY | 95 | PheGluSerIleIleHisGluPheAspProTrpPheAsnTrpPheAsnTrpHisHisLeu | 114 |
| Db | 350 | TCGAGAGCATCATCCAGAGTTCGACCGGTGGTAACTATAGATCAACACATCATCTT | 409 |
| QY | 115 | AlaSerHisGlyPheTyrCluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeu | 134 |
| Db | 410 | GCATCTCATGGATTCTATGAGTTCTAAATGGTTTGGTGAAGAGAGCATGGTACCCACTG | 469 |
| QY | 135 | GlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHis | 154 |
| Db | 470 | CGAAGAAATAGTGGTGGCACCGTTTACCAGGGTTGATGATAACAGCTGGCTTATTCAT | 529 |
| QY | 155 | TrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAcpValCysValPheLeuAla | 174 |
| Db | 530 | TGAATTTTAAATACATTGAACATAACAGTTACATAGAGATGGTGTGATTTCCCTTGA | 589 |
| QY | 175 | ProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsn | 194 |
| Db | 590 | CCAACTTTTAGCGGCTTACATCCATATCTAGTCTCTGCTAACTAGAGAACTGTGGAAC | 649 |
| QY | 195 | GlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArg | 214 |
| Db | 650 | CAAGGAGCAGGACTCTAGCTGCTTCTCATGCTATGCTACGACAGGATCATATCTCGG | 709 |
| QY | 215 | SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr | 234 |
| Db | 710 | TCAGTGGCGGATCCTTTGATATATAGAGCATTTGCCATTTTGGCTTCAGTTCACTTAC | 769 |
| QY | 235 | TyrLeuTrpValIleSerValIleThrGlySerValPheTrpThrMetCysCysLeu | 254 |
| Db | 770 | TACTTATGGGTAAAGTCTGTGAAGACCGGGTCTGGTTCGCAATGTGCTGCTGTG | 829 |
| QY | 255 | SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleAsnLeuIlePro | 274 |
| Db | 830 | TCATATTTCTACATGGTCTCTGCGTGGGAGGTTATGTTTCATCATCAACCTCATCCCT | 889 |
| QY | 275 | LeuHisValPheValLeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyr | 294 |
| Db | 890 | CTCCATGTGTTTGTGTGCTGTGATGAGAGGTACAGCAAGAGAGTCTACATAGCATAT | 949 |
| QY | 295 | SerThrPhe-TyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGln | 314 |
| Db | 950 | AGCAC-TTGTACATTTGGGTTTAAATATATCCATGCGAGATACCTTTTGGGATTTC | 1008 |
| QY | 314 | NProfileArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuGlnAlaTyr | 334 |
| Db | 1009 | GCCAAATCAAGCAAGGAGCAGCATGGAGCTGACAGGTCTTTGGCTGTGCAAGCTTA | 1068 |
| QY | 334 | rAlaPheLeuGlnTyrLeuArgAspArgLeuThrIleGlnPheGlnThrLeuPhePhe | 354 |
| Db | 1069 | CGCTTTTTCAGTATCTGAGAGACCGGTGACAAAAACAGAGTTCAGAGCCCTTTTCTT | 1128 |
| QY | 354 | eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrTh | 374 |
| Db | 1129 | TTTGGGTGTCTCACTAGCTGACGGCGCTGTGTTCTTAGTGTCTATCTATCTGACATAC | 1188 |

| | | | |
|----|------|--|------|
| QY | 374 | rGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLys | 394 |
| Db | 1189 | AGGTTATATTCCACCATGGAGTGGCAGGTTTATTTCACTATGGATACCTGGGTATGCAAA | 1248 |
| QY | 394 | sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpValSerPhe | 414 |
| Db | 1249 | AATACACATTCCTAATTTATGGCATCAGTGTCTGAACATCAGCCCTACGACATGGGTGCTTT | 1308 |
| QY | 414 | ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLys | 434 |
| Db | 1309 | CTTCTTTGATCTACATATTTCTGTATGATCTCTCCAGCAGGCTATGGTCTGTCATCAA | 1368 |
| QY | 434 | eAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl | 454 |
| Db | 1369 | AAATATCAACATGAAGAGATTTTTCGCTCTGTATGCGATCAGTGTGTGTCTTTC | 1428 |
| QY | 454 | aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaI | 474 |
| Db | 1429 | CGGAGTATGTTGGGCTGTGCTGACTCTGACCCCGGTCTCTGATGCTGTGCGCCAT | 1488 |
| QY | 474 | eAlaPheSerAsnValPheGluHisTyrIleGlyAspMetLeuArgGluAsnProPr | 494 |
| Db | 1489 | COCTTCTTCCAAATGTTTTTTCGACACTATTTGGGGATGACATGAAGAAAGGAAACCCACC | 1548 |
| QY | 494 | oValGluAspSerSerAspGluAspAspLysArgAsnGln-GlyAsnLeuTyrAspLysA | 514 |
| Db | 1549 | TGTGGAGGACAGCAGTGTATGAGGATGACAAAAGAAACCC-AGGAAACTTGTATGACAAGG | 1607 |
| QY | 514 | LaGlyValArgCysHisAla-ThrGluGlnGluLysThr-GluGluGlyLeuGlyPr | 533 |
| Db | 1608 | CAGGTAAAGTGAGGAGCATGT-GACAGAGCAGAGAA-ACCTGAAGAGGCTTGGGCC | 1665 |
| QY | 533 | oAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHi | 553 |
| Db | 1666 | CAAATCAAAAGCATTTGTACCATCTGCTGATGCTCATGCTCTGATGATGTTTCGCGTCCA | 1725 |
| QY | 553 | sCysThrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAs | 573 |
| Db | 1726 | CTGACGTGGGTCAAGCAAGCCTACTCAGTCCAGTGTGGTCTTTCCTCCCTACAA | 1785 |
| QY | 573 | nHisAspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgGln | 593 |
| Db | 1786 | TCATGATGGTACCAGGAATATATTAGATGATTTTAGAAGACGTACTTTTGGCTGAGACA | 1845 |
| QY | 593 | nAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGln | 613 |
| Db | 1846 | AAACACAGATGAACACGCCCGGTCATGCTGGTGGGACTACGGCTATCAGATTCTGG | 1905 |
| QY | 613 | yMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLe | 633 |
| Db | 1906 | CATGCCCAACAGGACCATCTGTTGGATAACAACACCTGGAAACAACAGCCACATCCCACT | 1965 |
| QY | 633 | uValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThr-LeuA | 653 |
| Db | 1966 | GCTCGGAAAGACTATGCTTCCAAATGAACGGCGCTATPAAATCATGAGGTCTCTTGG | 2024 |
| QY | 653 | spValAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleA | 673 |
| Db | 2025 | ATGTCGATTAATGTTGTTTATTTTCGAGGAGTGTGCTGCTATTCGGGGAGCATATCA | 2084 |
| QY | 673 | enLysPheLeuTrpMetValArgIleAlaGlnGlyGluHisProLysAspIleArgGln | 693 |
| Db | 2085 | ACAAGTTCTCTGGATGTGAGGATAGCTGAAGGGAGCATCCCAAGACATCCCGGAAG | 2144 |
| QY | 693 | er-AspTyrPheThrPro-GlnGlyGluPheArgValAspLysAlaGlySerProThrLe | 712 |
| Db | 2145 | G-TGACTATTTTAC-CCAGCAGGAGAGTTCAGAGTAGACAAAGCTGGGTCTCTACTCT | 2202 |
| QY | 712 | uLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPh | 732 |
| Db | 2203 | GTTAAATGCTCTATGATATAAATGCTACTACTAGATTGGAGAAATGCGAGCTAGATT | 2262 |
| QY | 732 | eArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLys | 752 |

QY 61 GlyLeuSerGlnProAlaGlyTyrGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 Db 208 GGGCTGTGCGACCGCGCTGGGTGGGAGTGGCTTCTCTCTCCATCCCTCTCTCTGGCC 267
 QY 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCGTCATCCGCTTCGAAAGCATCATCCAC 327
 QY 101 GluPheAspProTrpPheAsnTyrArgSerThrHisIleLeuAlaSerHisGlyPheTyr 120
 Db 328 GAGTTTCGACCGCGTGTAACTATATGATCAACATCATCTTGATCTCATGGTCTAT 387
 QY 121 GluPheLeuAsnTrpPheAspAlaArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 Db 388 GAATTTTAAATTTGGTTTTCATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGT 447
 QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160
 Db 448 ACTGTTTACCCAGGCTTATGATATACCGCTGGCTTATCATTTGGATTTTAAATACATG 507
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 Db 508 AACATAACTGTTTACATAAGAGACGTGTGTGTCTCTTGACCAACTTTTAGCGGCTT 567
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeu 200
 Db 568 ACATCTATATCTACTTTCTGCTTTACAAGAGACTTTTGGAAACCAAGAGCAGGACTTTTA 627
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
 Db 628 GCTGCTTGTATTGCTATTGTACAGGCTACATATCTCGTTCAGTAGCTGATCTCTT 687
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValIleSer 240
 Db 688 GATAATGAAGGCAATGCTATTTTGCACCTTCAGTTCACATATTTATGGGTAATCT 747
 QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetVal 260
 Db 748 GTAAACCTGGCTCAGTTTGTGGACAATGTGCTGCTTATCTATCTATATGCTC 807
 QY 261 SerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProIleHisValPheValLeu 280
 Db 808 TCTGCTTGGGGTGTATGTATTTATCATCATCTTATTCACATGCAATGATTTGTGTTG 867
 QY 281 LeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
 Db 868 TTACTCATGCAGATACAGCAAAAGAGCTTACATAGCATATAGCATCTTTCTACATTGT 927
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 Db 928 GGTTTAATATTAATCAATGCAATACCTTTTGTGGGATTCAGCAATCAGAAAGAGAA 987
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
 Db 988 CACATGGCAGCTGCGAGGTCTTTGCAATGCTGCAAGCTTATGCTTCTTCTGAGTATCTG 1047
 QY 341 ArgAspArgLeuThrIleGlnPheGlnLeuPhePheLeuGlyValSerLeuAla 360
 Db 1048 AGAGACCGGATTAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCAGTAGCT 1107
 QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrp 380
 Db 1108 GCAGGTGCTGTCTCTTAGTGTATCTATTTTGTATCTTATAGCTTACATGACCATG 1167
 QY 381 SerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaIleHisIleProIleIle 400
 Db 1168 ATGGCAGGTTTATCTATTTGGGATCTGGGTATGCAAAATACATCTCAATATTT 1227
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
 Db 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGTCTTCTCTTTGTATCTACATATT 1287

QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLeuAsnIleAspGluArg 440
 Db 1288 CTTGTATGTACCTTCCAGCAGCCCTTGGTTCTGCATCAAAATATCAACCATGAAGA 1347
 QY 441 ValPhe 442
 Db 1348 GTATTT 1353
 RESULT 4
 ADD94785
 ID ADD94785 standard; cDNA; 2710 BP.
 XX
 AC ADD94785;
 DT 29-JAN-2004 (first entry)
 XX
 DE Mouse SIMP cDNA sequence.
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA; cancer;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ss.
 XX
 OS Mus musculus.
 XX WO2003054008-A2.
 XX
 XX 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 XX 20-DEC-2001; 2001US-00028384.
 XX
 XX (COMP-) COMPATIGENE INC.
 XX
 XX Perreault C, McBride K;
 XX
 XX WPI; 2003-559122/52.
 XX P-FSDB; ADD94786.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 XX Disclosure; SEQ ID NO 3; 56pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the mouse SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;
 Alignment Scores:

Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGGTCATCGCTTCGAAAGCATCATCCAC 327
 QY 101 GluPheAspProTyrPheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyr 120
 Db 328 GAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTCATCTCATGGGTCTAT 367
 QY 121 GluPheLeuAsnTyrPheAspGluArgAlaTyrPyrProLeuGlyArgIleValGlyGly 140
 Db 398 GAATTTTAAATGTGTGTGAAGAGCATGGTATCCACTAGGAAAGATAGTAGGTGT 447
 QY 141 ThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTyrPyrLeuAsnThrLeu 160
 Db 448 ACTGTTTACCAGGCTTGATGATACCGCTGCTTATTCATTCGATTTTAAATACATG 507
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 Db 508 AACATAACTGTTTCAATAGAGAGCGTGTGTGTCTTCCACCACTTTTAGCGGCTT 567
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuThrPheGlnGlyAlaGlyLeuLeu 200
 Db 568 ACATCTATATCTACTTCTCTTACAGAGAACTTTTGAACCAAGGAGCAGACTTTTA 627
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPhe 220
 Db 628 GTGCTGTGTTTATGCTATGTACAGGCTACATATCTCGTCTAGTAGTGTATCTTT 687
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValLysSer 240
 Db 688 GATAATGAAGCATTTGCTATTTTGCACCTTCAGTTCACATCTATTTATGGTAAATCT 747
 QY 241 ValLysThrGlySerValPheThrThrMetCysCysLeuSerTyrPheTyrMetVal 260
 Db 748 GTAAATACCTGGGTCTAGTCTTTTGGCAATGTGCTGCTTATCTTCTATATATGCTC 807
 QY 261 SerAlaTyrGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 Db 808 TCTGCTGGGTGTGTATGTTATTTATCATCAATCTTATTCACATGCTGATTTTGTGTTG 867
 QY 281 LeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleVal 300
 Db 868 TTACTGATGACAGATACAGCAAAAGAGCTTACATAGATATACATCTTCTCATATGTTG 937
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 Db 928 GGTTTTAAATATATCAATGACATACCTTTTGTGGATTTCCAGCCAAATCAGAACAGTGAA 987
 QY 321 HisMetAlaAlaGlyValPheAlaLeuGlnAlaTyrAlaPheLeuGlnTyrLeu 340
 Db 989 CACATGGCAGCTGCAGGTGCTTTTGCATTTGCTGCAAGCTTATGCTTTCTTCAGTATCTG 1047
 QY 341 ArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 Db 1048 ACAGACCGATTAACAAACAAAGAGTTCCAGACCTTTTCTTTTGGGTGTATCACTAGCT 1107
 QY 361 AlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyr 380
 Db 1108 GCAGGTGCTGCTGCTTCTTAGTGTCTATCTATTGATCTATACAGTTTACATGACCATGG 1167
 QY 381 SerGlyArgPheTyrSerLeuThrPheThrGlyTyrAlaIleHisIleProIleIle 400
 Db 1168 AGTGGCAGGTTTATCTATTGTTGGGATCTGGGTATGCAAAATATACACATTTCAATATT 1227
 QY 401 AlaSerValSerGluHisGlnProThrThrThrThrValSerPhePheAspLeuHisIle 420
 Db 1228 GCATCAGTGTCTGAGATCAACCTACGACTTGGGTGCTTCTTCTTCTTGTATCTACATATT 1287
 QY 421 LeuValCysThrPheProAlaGlyLeuThrPheCysIleLysAsnIleAsnAspGluArg 440
 Db 1288 CTTGTATGTATCTTCCAGCAGGCTTTTGGTGTCTGCTCAAAATATCAACGATGAAGA 1347
 QY 441 ValPhe 442
 Db 1348 GTATTTT 1353

RESULT 3
 ADL30661
 ID ADL30661 standard; cDNA; 1664 BP.
 AC ADL30661;
 XX ADL30661;
 DT 20-MAY-2004 (first entry)
 DE Full length human cDNA clone SeqID 2694.
 XX human; medicine; signal transduction; glycoprotein; transcription;
 KW oligo-capping method; ss; gene.
 XX Homo sapiens.
 OS
 XX EPI396543-A2.
 PN
 XX 10-MAR-2004.
 PD
 XX 07-JUL-2000; 2003EP-00025638.
 PF
 XX 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183865.
 PR 07-JUL-2000; 2000EP-00114089.
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 PA
 XX Ota T. Nishikawa T. Isogai T. Hayaashi K. Ishii S. Kawai Y;
 PI Wakamatsu A. Sugiyama T. Nagai K. Kojima S. Otsuki T. Koga H;
 XX WPI; 2004-204755/20.
 DR P-PSDB; ADL30662.
 XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 PT length human cDNAs.
 PT
 PS Example 1; SEQ ID NO 2694; 1340pp; English.
 CC This invention relates to a novel primers useful for synthesizing full
 CC length cDNA molecules that encode human proteins. Specifically, it refers
 CC to secretory or membrane proteins that are potential therapeutic agents/
 CC target molecules in the field of medicine, and in particular genes
 CC encoding proteins that are associated with signal transduction.
 CC glycoproteins and transcription. The present invention describes a method
 CC for efficiently cloning a full length human cDNA from both the 5' and 3'
 CC ends using the oligo-capping method. This polynucleotide sequence is a
 CC full length human cDNA clone of the invention.
 XX
 SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 1664
 Score: 442.00 Matches: 442
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.51% Indels: 0
 DB: 12 Gaps: 0
 US-10-028-384-2 (1-826) x ADL30661 (1-1664)
 QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTyr 20
 Db 28 ATGGCGAGCGCTCGGCGCGGAGAGACAGCAAGTCTCACTCACTCGTCCCGCTGG 87
 QY 21 SerGlyLeuMetAlaLeuGlyArgHisGlyHisGlyProGlyAlaGlnCys 40
 Db 88 AGTGGCTCTATGCGCTTGGGAAACAGCCGACCGCCCGGGGCCAGTGC 147
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 Db 148 CGCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207

Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCGAGTAAAGTGAGAAACAT 1560
 Qy 521 AlaThrGluGlnGluThyGluGluGlyProAsnIleLysSerIleValThr 540
 Db 1561 GCACTGAACAGGAAAAAATCGAGAGGATAGGCCCTAATATPAAAGACATGTCACC 1620
 Qy 541 MetLeuMetLeuMetLeuMetMetPheAlaValHisCysThrTrpValThrSerAsn 560
 Db 1621 ATGTTGATGCTGATCTATTGATGATGTTGCTGCTCCACTGTACTCGGTCAAGACAAAT 1680
 Qy 561 AlaTySerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIle 580
 Db 1681 GCCTACTCTAGTCCCAAGTGTAGTCTGCTGCTCATACAAATCATGATGCCACCAAGATATC 1740
 Qy 581 LeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAlaArg 600
 Db 1741 TTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACCGA 1800
 Qy 601 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 620
 Db 1801 GTAATGCTTTGTTGGATTAATGGCTATCAGATAGCTGGATGGCTAATAGAACTAGCTTG 1860
 Qy 621 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 640
 Db 1861 GTGGATAATAACACCTGGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGTCTTCT 1920
 Qy 641 AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIle 660
 Db 1921 AATGAACAGACAGCTTAATAATCATGAGGACTCTAGATGATGATTAATGTTTGGTATT 1980
 Qy 661 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 680
 Db 1981 TTGGAGGGCTATTGGCTATTCTGATGATATCAACAAATTTCTCTGGATGGTAGG 2040
 Qy 681 IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly 700
 Db 2041 ATAGCTGAAGAGAGAACTCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGGA 2100
 Qy 701 GluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet 720
 Db 2101 GAATTCCTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTCCTTATGATATAAATG 2160
 Qy 721 SerTyrTyrArgPheGlyGluMetGluLeuAspPheArgThrProProGlyPheAspArg 740
 Db 2161 TCATCTACTACGATTGGAGAAATGGCGTGGATTTTCTACACCCCGAGGTTTGGACCGA 2220
 Qy 741 ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAlaPhe 760
 Db 2221 ACACGTAATGCTGAGATTGGAATAAGGACATTAATTAATCAACATTTGGAAGAAGCTTT 2280
 Qy 761 ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr 780
 Db 2281 ACATCAGAACACTGGCTTGTAGATATATTAAGATAAAGCCTGATATAACAGGGAGACA 2340
 Qy 781 LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys 800
 Db 2341 TTAGATCAACAACTCGAGTCACCAACATTTTCCCAAAACAGAAGTATTGTTCAAGAAG 2400
 Qy 801 ThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLysLys 820
 Db 2401 ACTACAAAGAGAGCGGTGGCTACTATTAATAAATAGCTGTTTAAAGAAAGCAAGAAA 2460
 Qy 821 IleSerLysLysThrVal 826
 Db 2461 ATATCTAAGAGACTGTT 2478

RESULT 2

AAK94164

ID AAK94164 standard; cDNA; 1664 BP.

XX AC

XX AAK94164;

XX DT

06-NOV-2001 (first entry)

XX Human full-length cDNA, SEQ ID NO: 2694.
 DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 OS EP1130094-A2.
 PN 05-SEP-2001.
 PD 07-JUL-2000; 2000EP-00114089.
 PP 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 DR P-PSDB; AAM93252.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 in genetic manipulation.
 PT Claim 8; SEQ ID NO 2694; 1380pp + Sequence Listing; English.
 XX

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

XX Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 1664
 Score: 442.00 Matches: 442
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 53.51% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-2 (1-826) x AAK94164 (1-1664)

Qy 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProTrp 20
 Db 28 ATGCGGAGCCCTCGGCCCGGAGACCAAGTCGTCCTCACTCGTCCCGGTG 87
 Qy 21 SerGlyLeuValAlaLeuGlyAsnSerArgHisGlyHisHisGlyProGlyAlaGlnCys 40
 Db 88 AGTGGCTCATGTCCTCGGAAACAGCCGGACAGCCACACGGGGCCCGGGCCAGTGC 147
 Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 Db 148 GCGCACAGCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 207
 Qy 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuLa 80
 Db 208 GGGCTGTGCGAGCGCGCTGGTGGCAGTCGCTCTCTCTCCTTCCATCTCTCTCTG 267
 Qy 81 TrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHis 100

XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 DR P-PSDB; ADD94784.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Claim 6; SEQ ID NO 1; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.

XX SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2481
 Score: 826.00 Matches: 826
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-2 (1-826) x ADD94783 (1-2481)

QY 1 MetAlaGluProSerAlaProGluSerLysHisLysSerSerLeuAsnSerSerProtrp 20
 DB 1 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCTCCTCAACTCGTCCCGTGG 60
 QY 21 SerGlyLeuMetAlaLeuGlyAsnSerArgHisGlyHisGlyProGlyAlaGlnCys 40
 DB 61 AGTGGCTCATGCGCTCGGNAACACCGGACCGCCACCGGGCCCGGGCCAGTGC 120
 QY 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAlaProAlaGlyLeuSerGly 60
 DB 121 GCGCAACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 QY 61 GlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrIleLeuPheLeuAla 80
 DB 181 GGGCTGCGCGCGGCTGGGCGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
 QY 81 TrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHis 100
 DB 241 TGGCTTGGCGGCTTCAGCTCGGCGCTTTCGCGGTCATCCGCTTCGAAGCATCATCCAC 300
 QY 101 GluPheAspProTrpPheAsnTyArgSerThrHisHisLeuAlaSerHisGlyPheTy 120
 DB 301 GAGTTGCGCGGCTTAACTATAGATCAACATCATCTTTCGATCTCATCGGTTCTAT 360
 QY 121 GluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArgIleValGlyGly 140
 DB 361 GAATTTTAAATGTTTTCATGAAGAGCATGGTATCCATCAGGAGGAATAGTAGTGGT 420
 QY 141 ThrValTyProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeu 160

DB 421 ACTGTTTACCAGGGTTGATGATAACCGCTGCGCTTATTCATGATTTTAAATACATTG 480
 QY 161 AsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGlyLeu 180
 DB 481 AACATAACTGTTTCAATAGAGACGATATGTGTGTTCTTGCACCAACTTTTAGCGGCTT 540
 QY 181 ThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValGlyLeuLeu 200
 DB 541 ACATCTATATCTACCTTCTCTACAGAGAACTTTGGACCAAGGAGGAGGAGCTTTTA 600
 QY 201 AlaAlaCysPheIleAlaIleValProGlyTyIleSerArgSerValAlaGlySerPhe 220
 DB 601 GTGTGCTGTGTTTATTCCTATTTACACAGGCTACATATCTCGTACGTAGTACGTGATCCTTT 660
 QY 221 AspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyIleLeuTrpValIleSer 240
 DB 661 GATAATGAAGGATGCTATTTTGCACTTCAGTTCACATACATATTTAAGGTAAATCT 720
 QY 241 ValIleThrGlySerValPheTrpThrMetCysCysLeuSerTyIlePheTyIleMetVal 260
 DB 721 GTAAAAAAGCTGGTCACTTTTGGACAAATGCTGCTGCTTATCTATCTATATGCTC 780
 QY 261 SerAlaTrpGlyGlyTyIleValPheIleIleAsnLeuIleProLeuHisValPheValLeu 280
 DB 781 TCTGCTTGGGTGGTGTATGATTTATCATCAATCTTATCCACTGCATGATTTGTGTG 840
 QY 281 LeuLeuMetGlnArgTyIleSerLysArgValTyIleAlaTyIleSerThrPheTyIleVal 300
 DB 841 TTACTGTATGACAGATACAGCAAAAGAGTCTACATAGCATATACACATCTTCTACATG 900
 QY 301 GlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGlu 320
 DB 901 GGTITATATATATCAATGACATACCTTTTGGGATTCAGCCCAATCAGAACAAAGTAA 960
 QY 321 HisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyIleAlaPheLeuGlnTyIle 340
 DB 961 CACATGGCAGCTGCGAGGTGCTTTCATGCTGCTGCAAGCTTATGCTTCTTGCATGATCTG 1020
 QY 341 ArgAspArgLeuThrLysGlnPheGlnThrLeuPhePheLeuGlyValSerLeuAla 360
 DB 1021 AGAGACCGATTAACAAACACAGAGTTCAGACCCCTTTCTTTGGGTGATCACTAGCT 1080
 QY 361 AlaGlyAlaValPheLeuSerValIleTyIleLeuThrTyIleAlaProTrp 380
 DB 1081 CGAGCTGCTGTGTTCTTCTAGTGTCTATTTGACTTATACAGGTTACATGACCAATGG 1140
 QY 381 SerGlyArgPheTyIleSerLeuTrpAspThrGlyTyIleAlaLysIleHisIleProIle 400
 DB 1141 AGTGGCAGGTTTATTCATTTGGGATCTGGGTATGCAAAAAATATCAACAGATGAAGA 1200
 QY 401 AlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHisIle 420
 DB 1201 GCATCAGTGTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTATCTACATATT 1260
 QY 421 LeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArg 440
 DB 1261 CTTGTATGTACTTCCCGCAGCGCTTTGGTTCCTGCATCAAAATATCAACAGATGAAGA 1320
 QY 441 ValPheValAlaLeuTyIleAlaIleSerAlaValTyIlePheAlaGlyValMetValArgLeu 460
 DB 1321 GTATTGCTGCTCTATATGCAATCAGTGTCTCTCTTGTCTGAGTGTGTTGGGAGTGC 1380
 QY 461 MetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPhe 480
 DB 1381 ATGTTGACTTTGACTCCAGTCGTGTGTATGCTGCTGCTGCAATTCCTTTTCAATGTTTT 1440
 QY 481 GluHisTyIleGlyAspAspMetLysArgGluAsnProProValGluAspSerSerAsp 500
 DB 1441 GAGCATTATTGGGGATGACATGAAGAGGAAAAATCCACCTGTGGAGCAGCAGCTGAT 1500
 QY 501 GluAspAspLysArgAsnGlnGlyAsnLeuTyIleAspLysAlaGlyLysValArgLysHis 520

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57; Search time 951.69 Seconds

(without alignments)

4556.130 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 826

Sequence: 1 MAEPSAPESKHSKSLNSGSPW.....GYIKNKLVPKKGKISKKTIV 826

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+P2n.model -DEV=xlh
-Q/cn2_1/USPTO.spool/US10028384/runat_14122004.131645.6790/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -QMT=fastcap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384 @CGN 1.1.1986 @runat 14122004.131645.6790 -NCPU=3
-NO WMAP -LARGSEQ=NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

N Geneseq 23Sep04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2002bs: *
8: Geneseqn2003as: *
9: Geneseqn2003bs: *
10: Geneseqn2003cs: *
11: Geneseqn2003ds: *
12: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| 1 | 826 | 100.0 | 2481 | 10 | ADD94783 | ADD94783 Human sim |
| 2 | 442 | 53.5 | 1664 | 4 | AAK94164 | AAK94164 Human ful |
| 3 | 442 | 53.5 | 1664 | 12 | ADL30661 | ADL30661 Full leng |
| 4 | 365 | 44.2 | 2710 | 10 | ADD94785 | ADD94785 Mouse SIM |
| 5 | 351 | 42.5 | 1209 | 4 | AAH33264 | AAH33264 Human col |
| 6 | 351 | 42.5 | 1209 | 6 | ABL9850 | ABL9850 Human pol |

| | | | | | | |
|----|-----|------|------|----|----------|--------------------|
| 7 | 351 | 42.5 | 1543 | 2 | AAK95055 | AAK95055 Human sec |
| 8 | 351 | 42.5 | 1543 | 8 | ACD18981 | ACD18981 Novel hum |
| 9 | 351 | 42.5 | 1543 | 12 | ADG78372 | ADG78372 Human sec |
| 10 | 351 | 42.5 | 1543 | 12 | ADN60663 | ADN60663 Human sec |
| 11 | 323 | 39.1 | 2546 | 5 | AAV44866 | AAV44866 Clone CTS |
| 12 | 323 | 39.1 | 2546 | 5 | AAV98463 | AAV98463 Human cDN |
| 13 | 317 | 38.4 | 2510 | 5 | AAF93772 | AAF93772 Human cDN |
| 14 | 297 | 36.0 | 2537 | 4 | AAO08289 | AAO08289 Human sec |
| 15 | 286 | 34.6 | 2284 | 4 | AAH18021 | AAH18021 Human cDN |
| 16 | 217 | 26.3 | 1114 | 4 | AAH99794 | AAH99794 Human pro |
| 17 | 196 | 23.7 | 2547 | 4 | AAO08315 | AAO08315 Human sec |
| 18 | 196 | 23.7 | 2660 | 6 | ABQ54750 | ABQ54750 Human ova |
| 19 | 168 | 20.3 | 507 | 3 | AAO01967 | AAO01967 Human sec |
| 20 | 166 | 20.2 | 787 | 4 | AAH07526 | AAH07526 Human cDN |
| 21 | 166 | 20.1 | 500 | 6 | ABV88332 | ABV88332 Human col |
| 22 | 145 | 17.6 | 764 | 5 | AAF93958 | AAF93958 Primer sp |
| 23 | 143 | 17.3 | 433 | 5 | ABL37128 | ABL37128 Human col |
| 24 | 133 | 16.1 | 558 | 5 | AAF93612 | AAF93612 Umbilical |
| 25 | 119 | 14.4 | 387 | 4 | ABA08397 | ABA08397 Human sec |
| 26 | 117 | 14.2 | 616 | 3 | AAZ80248 | AAZ80248 Human col |
| 27 | 110 | 13.3 | 483 | 3 | AAA43486 | AAA43486 Mouse sec |
| 28 | 108 | 13.1 | 485 | 9 | ACH24071 | ACH24071 Human adu |
| 29 | 107 | 13.0 | 349 | 2 | AAV89737 | AAV89737 EST clone |
| 30 | 94 | 11.4 | 313 | 8 | ABZ20328 | ABZ20328 Group III |
| 31 | 82 | 9.9 | 250 | 3 | AAO01440 | AAO01440 Human sec |
| 32 | 78 | 9.4 | 245 | 3 | AAO17015 | AAO17015 Human sec |
| 33 | 78 | 9.4 | 247 | 2 | AAK40252 | AAK40252 Human sec |
| 34 | 77 | 9.3 | 281 | 3 | AAA45221 | AAA45221 Human sec |
| 35 | 62 | 7.5 | 388 | 5 | AAF66328 | AAF66328 Novel hum |
| 36 | 54 | 6.5 | 474 | 5 | AAF93633 | AAF93633 Umbilical |
| 37 | 51 | 6.2 | 393 | 4 | ABA08547 | ABA08547 Human sec |
| 38 | 50 | 6.1 | 503 | 4 | AAK91644 | AAK91644 Human cDN |
| 39 | 50 | 6.1 | 503 | 4 | AAK93577 | AAK93577 Human cDN |
| 40 | 50 | 6.1 | 503 | 12 | ADL30004 | ADL30004 3' end of |
| 41 | 50 | 6.1 | 503 | 12 | ADL28071 | ADL28071 5' end of |
| 42 | 44 | 5.3 | 737 | 4 | AAI96771 | AAI96771 Human neu |
| 43 | 42 | 5.1 | 2417 | 10 | ADD94789 | ADD94789 Drosophil |
| 44 | 42 | 5.1 | 2699 | 4 | ABL13247 | ABL13247 Drosophil |
| 45 | 42 | 5.1 | 4922 | 4 | ABL13246 | ABL13246 Drosophil |

ALIGNMENTS

RESULT 1

ADD94783
ID ADD94783 standard; cDNA; 2481 bp.

XX AC ADD94783;

XX DT 29-JAN-2004 (first entry)

XX DE Human SIMP cDNA sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; VHC;
XX KW major histocompatibility complex; human leukocyte antigen; HLA;
XX KW cystostatic; immunosuppressive; antitumor therapy; Gene therapy; cancer;
XX KW lung cancer; intestine cancer; sarcoma; prostate cancer;
XX KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
XX KW hematological cancer; immune response; lymphoid cell proliferation;
XX KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
XX KW Gene; ss.

XX OS Homo sapiens.

XX FN WO2003054008-A2.

XX PD 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CA001967.

XX PR 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers

source

1..798

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30659262"
 /tissue_type="Upper Head"
 /dev_stage="embryo 9.5 - 10.5 dpc"
 /lab_host="pDH10B (T1 phage resistant)"
 /clone_lib="NTH_BMAP_H00"

/note="Organ: Upper Head; Vector: pYX-Asc; Site: 1: EcoR I; Site: 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

Query Match 27.2%; Score 674.6; DB 7; Length 798;
 Best Local Similarity 91.0%; Fred. No. 3.8e-168;
 Matches 727; Conservative 0; Mismatches 71; Indels 1; Gaps 1;

QY 500 GAGACGTATGTGTCTCTTGCACCACTTTTAGCGGCTTACATCTATATCTACTTTCC 559
 DB 1 GAGATGTGTGTATTCCTTGCACCACTTTTAGCGGCTTACATCTATATCTACTTTCC 60

QY 560 TCGTTACAGAGCACTTTGGACCAAGGAGGAGGAGCTTTAGCTGCTGTTTATTGCTA 619
 DB 61 TCGTAACTAGAGACTTGGACCAAGGAGGAGGAGCTTTAGCTGCTGTTTATTGCTA 120

QY 620 TTGTACCAAGGCTACATATCTCGGTGAGTGGATCCTTTGATAATGAAGGCAATTGCTA 679
 DB 121 TCGTACCAAGGCTACATATCTCGGTGAGTGGATCCTTTGATAATGAAGGCAATTGCTA 180

QY 680 TTTTTCGCTTCAGTTACATCTATTTATGGGTAATCTGTAAAGCTGGGCTGAGTTT 739
 DB 181 TTTTTCGCTTCAGTTACATCTATTTATGGGTAATCTGTAAAGCTGGGCTGAGTTT 240

QY 740 TTGGACAATGTGCTGCTTATCTCTATTTCTATATGGTCTCTGCTTGGGTTGTTATG 799
 DB 241 TCTGGACAATGTGCTGCTTATCTCTATTTCTATATGGTCTCTGCTTGGGTTGTTATG 300

QY 800 TATTTATCATCAATCTTATTCACCTGCAATGTTTGTGTTTACTGATGCAGAGATACA 859
 DB 301 TGTTCATCATCAACCTCATCCCTCTCCATGTGTTTGTGTTGCTGCTGATGCAGAGTACA 360

QY 860 GCAGAAGCTCTACATAGCATATAGCACTTTCTACATGTGGGTTTAATATCAATGC 919
 DB 361 TCAAGAGAGTCTACATAGCATATAGCACTTTCTACATGTGGGTTTAATATCAATGC 420

QY 920 AGATACCTTTTGTGGATTCAGCCCAATCAGAACAAAGTGAACATATGGCAGCTCCAGGTG 979
 DB 421 AGATACCTTTTGTGGATTCAGCCCAATCAGAACAAAGTGAACATATGGCAGCTCCAGGTG 480

QY 980 TCTTTGCATGTCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTACAAAAAC 1039
 DB 481 TCTTTGCATGTCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTACAAAAAC 540

QY 1040 AAGAGTTCAGACCCCTTTCTTTTGGGTGATCATAGCTTGCAGGTGCTGTGTTCTTA 1099
 DB 541 AAGAGTTCAGACCCCTTTCTTTTGGGTGATCATAGCTTGCAGGTGCTGTGTTCTTA 600

QY 1100 GTGTCTATCTATTTGACTTATACAGGTTATACATTTGCCACCATGGAGTGGCAGGTTTATTCAT 1159
 DB 601 GTGTCTATCTATTTGACTTATACAGGTTATACATTTGCCACCATGGAGTGGCAGGTTTATTCAT 660

QY 1160 TGTGGGATCTGGGTATGCARAAATACACATTCGAATTTTGCATCAGTGTCTGAGCATC 1219
 DB 661 TATGGGATCTGGGTATGC-AAAATACACATTTCCAAATTTANTGCAATCAGTGTCTGAGCATC 719

QY 1220 AACTTACGACTTGGGTGCTTTCTTTTGTGATCTACATATTTCTTGTATGTACCTTCCAG 1279
 DB 720 AGCCTACGACATGGGTGCTTTCTTTTGTGATCTACATATTTCTTGTATGTACCTTCCAG 779

QY 1280 CAGGCTTTGGTTCTGCAAT 1298
 DB 780 CANGGCTATGGTTCTGCAAT 798

Search completed: December 15, 2004, 11:09:42
 Job time : 7555.8 secs

| | | | |
|------|----|--|------|
| 1372 | QY | GTGGCACTGATGTTTGACTTTGACCTCCAGTCGGTGTGTATGCTGTCTGTCGAATTCCTTTTCA | 1431 |
| 505 | Db | GTGGCGCTGATGCTGACTCTGACCCCGGTGCTGTCATGCTGTGCGCCGTCGCCCTTCTCC | 564 |
| 1432 | QY | AATGTTTTTGGAGCACTATTTGGGGGGATGACATGAAAGGGAATTCACCTGTGGAGGAC | 1491 |
| 565 | Db | AATGTTTTTGGAGCACTATNTGGGGGATGACATGAAGAAGGAAAAACCHACCTGTGGAGGAC | 634 |
| 1492 | QY | AGCAGTGATGAGGATGACAAAAAGAAACCAAGGAAATTTGTATGATAAGGCAGGTAAGAAGTG | 1551 |
| 625 | Db | AGCAGTGATGAGGATGACAAAAGAGACCCAGGAACTTGTATGACAAGGCAGGTAAGAAGTG | 684 |
| 1552 | QY | AGGAAACATGCAACTGGAACAGGAAAAAACTGAAGAGGGATTAGGCCCTAATAATAAAAGC | 1611 |
| 685 | Db | AGGAAGCATGTGTACAGAGGCAAGAGAAAACCTGAAGAGGGCTTTGGGCCCAAGATCAAAAGC | 744 |
| 1612 | QY | ATTGTCACCATGTTGATGCTGATGCTATTATGATGATGTTTGCTGTCCACTGATCACTGGGTC | 1671 |
| 745 | Db | ATGTGTACCATCTGATGCTGCTCTCTGATGATGTACGGGGAGCACTGCACGTGGGTC | 804 |
| 1672 | QY | ACAAGCAATGCTACTCTAGTCCAAGTG-TAGTCCCTGGCCTCATCAATCATGATGGCAC | 1730 |
| 805 | Db | ACAAGGCAAGCCTACTCCAGTCCAGTGTGTGGTCTGACCTCTACATCATGATGATG | 864 |
| 1731 | QY | CAG 1733 | |
| 865 | Db | CAG 867 | |

CF995035 861 bp mRNA linear EST 25-NOV-2003
 LOCUS
 DEFINITION
 AGENCEOCT 15622198 NIH MGC 147 Homo sapiens CDNA clone
 IMAGE:30528919 5', mRNA sequence.

primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NIHRI, National Institutes of Health). Note: This is a NIH/MGC library."

| ORIGIN | Query Match | 27.6%; | Score 685.2; | DB 7; | Length 861; |
|--------|-----------------------|---|---------------------|---------------|-------------------|
| | Best Local Similarity | 98.7%; | Prod. No. 5.8e-171; | | |
| | Matches 701; | Conservative | 0; | Mismatches 8; | Indels 1; Gaps 1; |
| QY | 1691 | GCCTACTCTAGTCCAAAGTGTAGTCTCTGGCCCTCAVACAATCATGATGCGACACGGAATATC | 1740 | | |
| DB | 27 | GCCTACTCTAGTCCAAAGTGTAGTCTCTGGCCCTCATACATCATGATGCGACACGGAATATC | 86 | | |
| QY | 1741 | TTAGATGATTTTATAGAAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGACCGA | 1800 | | |
| DB | 87 | TTAGATGATTTTATAGAAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGACCGA | 146 | | |
| QY | 1801 | GTAATGCTCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTTGG | 1860 | | |
| DB | 147 | GTATGTCCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTTGG | 206 | | |
| QY | 1861 | GTGGATTAATACACCTGGAAATTAACAGCCACATAGACATCTGTGTGGAAAGCTATGCTCTCT | 1920 | | |
| DB | 207 | GTGGATTAATTAACACCTGGAAATTAACAGCCACATAGACATCTGTGTGGAAAGCTATGCTCT | 266 | | |
| QY | 1921 | AATGAACACAGCCCTATAAAATCATAGGAGCTCTAGATGTAGATTATGTTTGGTATTAT | 1980 | | |
| DB | 267 | AATGAACACAGCCCTATAAAATCATAGGAGCTCTAGATGTAGATTATGTTTGGTATTAT | 326 | | |
| QY | 1981 | TTTGGAGGGTTATTGGCTATTCTGGTGAATGATATCAAAATTTCTCTGGATGGTTAGG | 2040 | | |
| DB | 327 | TTTGGAGGGTTATTGGCTATTCTGGTGAATGATATCAAAATTTCTCTGGATGGTTAGG | 386 | | |
| QY | 2041 | ATAGCTCAAGAGAGAAATCCCAAGACATTCGGGAAAGTGACTATTTTATCCCCACAGGGA | 2100 | | |
| DB | 387 | ATAGCTGAAGAGAGAAATCCCAAGACATTCGGGAAAGTGACTATTTTATCCCCACAGGGA | 446 | | |
| QY | 2101 | GAATTCGGTGTAGACAAAGACAGATCCCGCTACTTTTGGTAATTCGCTTATGTTATAAAATG | 2160 | | |
| DB | 447 | GAATTCGGTGTAGACAAAGACAGATCCCGCTACTTTTGGTAATTCGCTTATGTTATAAAATG | 506 | | |
| QY | 2161 | TCATACTACAGATTTTCGAGAAATGACGTGGATTTTCGTACACCCCGAGGTTTGACCGA | 2220 | | |
| DB | 507 | TCATACTACAGATTTTCGAGAAATGACGTGGATTTTCGTACACCCCGAGGTTTGACCGA | 566 | | |
| QY | 2221 | ACACGTAATCCTCAGATTGGAAATTAAGGACATTAAATTCAAACATTTTGGGAAGAGCCTTT | 2280 | | |
| DB | 567 | ACACGTAATCCTCAGATTGGAAATTAAGGACATTAAATTCAAACATTTTGGGAAGAGCCTTT | 626 | | |
| QY | 2281 | ACATTCAGAACACCTGGC-TTGTGTAGGATATATAAGTAAAGACCTCATTAACAGGGAGAC | 2339 | | |
| DB | 627 | ACATTCAGAACACCTGGCTTTGTGTAGGATATATAAGTAAAGACCTCATTAACAGGGAGAC | 686 | | |
| QY | 2340 | ATTAGATCACAACACCTCGAGTCACCAACATTTTCCCAAAACAGAAATTT | 2389 | | |
| DB | 687 | ATTAGATCACAACACCTCGAGTCACCAACATTTTCCCAAAACAGAAATTT | 736 | | |

| | | | | |
|------------|--|--------|--------|-----------------|
| RESULT 15 | CN527592 | 798 bp | linear | EST 29-APR-2004 |
| LOCUS | CN527592 | | mRNA | |
| DEFINITION | IMAGE3065-cpe-b-07-0-UI.r1 NIH_BMAP_HQ00 Mus musculus cdna clone | | | |
| ACCESSION | IMAGE3065262 5', mRNA sequence. | | | |
| VERSION | CN527592 | | | |
| KEYWORDS | CN527592.1 GI:46855748 | | | |
| SOURCE | EST. | | | |
| ORGANISM | Mus musculus (house mouse) | | | |
| REFERENCE | 1 (bases 1 to 798) | | | |
| | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | |
| | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | | | |

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 823)
 REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs@mail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Cloning Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..823

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30612197"

/tissue_type="whole eye"

/dev_stage="embryo 12.5, 13.5, 14.5 dpc"

/lab_host="pH10B (T1 phage resistant)"

/clone_lib="NIH-BMAP_H00"

/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTTTCAAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

FEATURES

source

ORIGIN

Query Match 28.0%; Score 695.8; DB 7; Length 823;
 Best Local Similarity 91.1%; Pred. No. 8.5e-174;
 Matches 750; Conservative 0; Mismatches 72; Indels 1; Gaps 1;
 593 GACTTTAGCTGCTTTTATTGCTATTGTACAGGCTACATCTCGGTCACTAGCTG 652
 2 GCCTTCTAGCTCTTCTGCTTCAATGCTATCGTACAGGCTACATCTCGGTCACTAGCTG 61
 653 GATCCTTTGATTAATGAAGGCTATCTTTTGTCACTTCAGTTCACATCTTTATGGG 712
 62 GATCCTTTGATTAATGAAGGCTATGCTTTTGGGCTTCAGTTCACATCTTTATGGG 121
 713 TAAATCTGTAAAACTGGGTCAGTTTGTGGCAATGCTGCTGCTATCTCTTTCT 772
 122 TAAAGCTGTGAAGACCGGCTGCTGTTCTGGCAATGCTGCTGCTTGTCTATTTCT 181
 773 ATATGCTCTCTGCTGGGCTGCTGTTATTTATTCATCAATCTTATTCACATGAT 832
 182 ACATGCTCTCTGCTGGGAGGTTATGTTTCATCAACCTCATCTCTCCATGTT 241
 833 TTGTGTTGTTACTGATCAGAGATACAGCAAAAGAGTCTCATAGCATATAGCATTCT 892
 242 TTGTGTTGTTGCTGATCAGAGGTTACAGCAAGAGTCTCATAGCATATAGCATTCT 301
 893 ACATTTGTTGTTTATATATATATCAATGTCAGATACCTTTTGGGATTCAGCAATCAGAA 952
 302 ACATTTGTTGTTTATATATATATCAATGTCAGATACCTTTTGGGATTCAGCAATCAGAA 361

QY 953 CAAGTGAACACATGCGAGCTGCGAGGTGCTTTTGGCATTTGCTGCAAGCTTATGCTTTCTTGC 1012
 DB 362 CAAGCGAGACATGGCAGCTGCGAGGTGCTTTTGGCATTTGCTGCAAGCTTATGCTTTTTCG 421
 QY 1013 AGTATCTGAGAGACCGATTAAACAAACAGAGTTCACAGACCTTTTCTTTTGGGTGAT 1072
 DB 422 AGTATCTGAGAGACCGATTAAACAAACAGAGTTCACAGACCTTTTCTTTTGGGTGAT 481
 QY 1073 CACTAGTCTGAGGTGCTGTTTCTTCTTGTGTCATCTATTGACTTATACAGGTTACATTTG 1132
 DB 482 CACTAGTCTGAGGTGCTGTTTCTTCTTGTGTCATCTATTGACTTATACAGGTTATTTG 541
 QY 1133 CACATGAGTGGCAGGTTTATTTATTTGAGGATCTGGGTATGCAAAATACATTC 1192
 DB 542 CACATGAGTGGCAGGTTTATTTATTTGAGGATCTGGGTATGCAAAATACATTC 601
 QY 1193 CAATTATTGATCAGTGTCTGAGCATCAACCTTACAGTGGGTGCTTTTCTTTTCTTCTGATC 1252
 DB 602 CAATTATTGATCAGTGTCTGAGCATCAACCTTACAGTGGGTGCTTTTCTTTTCTTCTGATC 661
 QY 1253 TACATATTCTTGTATGCTACCTTCCAGCAGGCTTTGGTCTGTCATCAAAATACATTC 1312
 DB 662 TACATATTCTTGTATGCTACCTTCCAGCAGGCTTATGGTCTGTCATCAAAATACATTC 720
 QY 1313 ATGAAAGATGATTGTTGCTCTATATGCAATCAGTGTCTTCTTCTTCTGAGTGTG 1372
 DB 721 ATGAAAGATGATTGTTGCTCTTGTATGCAATCAGTGTCTTCTTCTTCTGAGTGTG 780
 QY 1373 TGGGACTGATTTGACTTTGACTCCAGTCTGCTGTCATCTCTC 1415
 DB 781 TGGGCTGATGCTGACTCTGACCCCGCTCGTCTGCACTGCTC 823

RESULT 12
 BG573592 778 bp mRNA linear EST 10-APR-2001
 LOCUS 602594608F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:472264 5',
 DEFINITION mRNA sequence.
 BG573592
 ACCSSION BG573592
 VERSION BG573592.1 GI:13581245
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 778)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM1579 row: 0 column: 01
 High quality sequence stop: 730.

FEATURES

source

1..778

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:472264"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 79"

/note="Organ: placenta; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI (ggcgctatggcc); 5' and 3' adaptors were used in cloning as follows: 5' -CAGGCGCATTAAGGCC-3'

| | | | |
|----|------|---|------|
| Qy | 1794 | TGCACGAGTAATGCTTGGTGGGATATATGGCTATCAGATAGCTGGAATGGCTAATAGAAC | 1853 |
| Db | 361 | TGCACGAGTAATGCTTGGTGGGATATATGGCTATCAGATAGCTGGAATGGCTAATAGAAC | 420 |
| Qy | 1854 | TACGTTGGTGATATAATACACCTCGGAATACACGCCACATAGCACCTGGTGGGAAAAAGCTAT | 1913 |
| Db | 421 | TACGTTGGTGATATAATACACCTCGGAATACACGCCACATAGCACCTGGTGGGAAAAAGCTAT | 480 |
| Qy | 1914 | GTCCTCTAAATGAAACACGACGCTATAAATCATAGGACTCTAGATGTAGATATATGTTTT | 1973 |
| Db | 481 | GTCCTCTAAATGAAACACGACGCTATAAATCATAGGACTCTAGATGTAGATATATGTTTT | 540 |
| Qy | 1974 | GGTTATTTTGGAGGGGTTATTTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGAT | 2033 |
| Db | 541 | GGTTATTTTGGAGGGGTTATTTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGAT | 600 |
| Qy | 2034 | GGTTAGGATAGCTCAAGGAGCAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCC | 2093 |
| Db | 601 | GGTTAGGATAGCTCAAGGAGCAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCC | 660 |
| Qy | 2094 | ACAGGGAGATTCGGTGTAGACAAAGCAGATCCCGCTACTTTGTGGAATGCCCTATGTGA | 2153 |
| Db | 661 | ACAGGGAGATTCGGTGTAGACAAAGCAGATCCCGCTACTTTGTGGAATGCCCTATGTGA | 720 |
| Qy | 2154 | TAAAAATGTCATACTA 2168 | |
| Db | 721 | TAAAAATGTCATACTA 735 | |

```

RESULT 10
BQ947308
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BQ947308 879 bp mRNA linear EST 21-AUG-2002
AGENCOURT 8955689 NCI CGAP Mam2 Mus musculus cDNA clone
IMAGE:6439669 5', mRNA sequence.
BQ947308
BQ947308.1 GI:22362786
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (Bases 1 to 879)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13960 row: a column: 14
High quality sequence stop: 675.
Location/Qualifiers
1..879
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6439669"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam2"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Salt;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

```

| Query Match | 28.1%; | Score 698.4; | DB 5; | Length 879; |
|-----------------------|--------------|--|----------------|-------------|
| Best Local Similarity | 88.5%; | Pred. No. 1.8e-174; | | |
| Matches 768; | Conservative | 0; | Mismatches 98; | Indels 2; |
| Gaps | 1; | | | |
| Qy | 1040 | AAGAGTTCAGACCCCTTTTCTTTTGGGTGATCACTAGCTGCAGTGCTGTGTTCCCTTA | 1099 | |
| Db | 12 | AGGAGTTCAGACCCCTTTTCTTTTGGGTGCTCACTAGCTGCAGGCCGTGTGTTCCCTTA | 71 | |
| Qy | 1100 | GTGTCATCTATTGACTTATACAGGTTACATTGCACCATGGAGTGGCAGGTTTATTTCAT | 1159 | |
| Db | 72 | GTGTCATCTATCTGCACATATACAGGTTATATTGCACCATGGAGTGGCAGGTTTATTTCAC | 131 | |
| Qy | 1160 | TGTGGGATACTGGGTATCGAAAAATACACATTCGAAATTTTGCATCAGTGCTCGAGATC | 1219 | |
| Db | 132 | TATGGGATACTGGGTATCGAAAAATACACATTCGAAATTTTGCATCAGTGCTCGAACATC | 191 | |
| Qy | 1220 | AACCTACGACTTTGGGTGTCTTTCTTTTGATCTACATATCTTGATGTACCTTCCGAG | 1279 | |
| Db | 192 | AGCCTACGACATGGGTGTCTTTCTTTTGATCTACATATCTTGATGTACCTTCCGAG | 251 | |
| Qy | 1280 | CAGGCTTTGGTCTGTCATCAAAAATATCACGATGGAAGATATTGTTGCTCTATATG | 1339 | |
| Db | 252 | CAGGCTATGGTCTGTCATCAAAAATATCACGATGGAAGATGTTTGCCTCTGTATG | 311 | |
| Qy | 1340 | CAATCAGTGTCTGTACTTTGCTGGAGTGATGGTGGCATGATGTGACTTTCGACTCCAG | 1399 | |
| Db | 312 | CGATCAGTGTCTGTACTTTGCTGGAGTGATGGTGGCATGATGTGACTTTCGACCCGG | 371 | |
| Qy | 1400 | TGCTGTGATGCTGTCTGCMAATTTGCTTTTCAAATGTTTTTGGACACATTTTGGGGATG | 1459 | |
| Db | 372 | TGCTGTGATGCTGTCTGGCGCTGCGCTTCTCCAATGTTTTTGGACACATTTTGGGGATG | 431 | |
| Qy | 1460 | ACATGAAAAGGAAAAATCCACTGTGGAGGACAGCAGTGCATGAGATGACAAAAGAAAACC | 1519 | |
| Db | 432 | ACATGAAAAGGAAAAACCCACTGTGGAGGACAGCAGTGCATGAGATGACAAAAGAAAACC | 491 | |
| Qy | 1520 | AGGGAATTTGTATGATTAAGCAGGTAAAGTGGAAAAATGCAACTGAAACNGGAAAAAA | 1579 | |
| Db | 492 | CAGGAAACTTTGATGACAGGCAGGTAAAGTGGAAAAATGCAACTGAAACNGGAAAAAA | 551 | |
| Qy | 1580 | CTGAAGAGGGATTTAGGCCCTTAATAAAAAAGCATTTGCACATGTTGATGCTGATGCTAT | 1639 | |
| Db | 552 | CTGAAGAGGGCTTTGGGCCCAACATCAAAAGCATTTGTCACCATGCTGATGCTCATCTCC | 611 | |
| Qy | 1640 | TGATGATGTTTCTGTCTCACTGTGACTCGGTGTCACAGCAATGCCCTACTCTCTAGTCCAAGTG | 1699 | |
| Db | 612 | TGATGATGTTTCTGTCTCACTGTGACTCGGTGTCACAGCAATGCCCTACTCTCTAGTCCAAGTG | 671 | |
| Qy | 1700 | TAGTCTCTGGCCCTCATACAATCATGATGCACACAGGAATATCTTAGATCATTTTAGAGAAG | 1759 | |
| Db | 672 | TGGTCTTTGCCCTCTCAATCATGATGTTACCGAGNATATATTAGATGATTTTAGAGAAG | 731 | |
| Qy | 1760 | CTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAC--GAGTAAATGCTTGGTGGGA | 1817 | |
| Db | 732 | CGTACTTTTGGCTGAGCAAAACAGGATGAACACGCCCGCGGTCATGCTCGTGGGTGGGA | 791 | |
| Qy | 1818 | TTATGGCTATCAGATGCTGGAATGGCTTAATAGAACTACGTTTGGTGGATATAACACCTG | 1877 | |
| Db | 792 | CTACGGCTATCAGATGCTTGGCAGGGCCACACAGGACCACTCTCTGTGGATTAACACCTG | 851 | |
| Qy | 1878 | GAATTAACGCCACATAGCACTGGTGGGA | 1905 | |
| Db | 852 | GGACAAACAGGCACATCGCACGGGTCCGA | 879 | |

[illegible]

primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTATTGAAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."

ORIGIN

| | | | | |
|-----------------------|--------|--|-------|----------------------------------|
| Query Match | 29.5%; | Score 731.2; | DB 7; | Length 852; |
| Best Local Similarity | 91.0%; | Pred. No. 3.1e-183; | | |
| Mismatches | 775; | Conservative | 0; | Mismatches 77; Indels 0; Gaps 0; |
| Qy | 578 | GGNACCAAGGACGAGGACTTTTACGCTGCTGTTTATTTGCTATTGTACGAGCTACATAT | 637 | |
| Db | 1 | GGAAACCAAGGACGAGGACTTTACGCTGCTGTTTATTTGCTATTGTACGAGGCTACATAT | 60 | |
| Qy | 638 | CTCGGTCAAGTACGATCCCTTTGATTAATGAAGCATTGCTATTATTTGCACTTCAGTTCA | 697 | |
| Db | 61 | CTCGGTCAAGTACGATCCCTTTGATTAATGAAGCATTGCTATTATTTGCACTTCAGTTCA | 120 | |
| Qy | 698 | CATACATTTTGGGTAAATCTGTAAACTGGGTCAGTTTTCGCAATGCTGCT | 757 | |
| Db | 121 | CTTACTATTGGGTAAAGTCTGTGAAGACCGGGTCTGTGTTCTGGCAATGCTGCT | 180 | |
| Qy | 758 | GCTTATCTCTATTCTATATGCTCTGCTTGGGTGGTATGTTATTTATCATCAATCTTA | 817 | |
| Db | 181 | GCTTGTGATATTCTACATGCTCTCGTGGGAGGTTATGTGTTTCATCATCAACTCA | 240 | |
| Qy | 818 | TTCCACTGCTGATTTTGTGTTTACTGATGACAGATACACAAAGCTCATATG | 877 | |
| Db | 241 | TCCTCTCTCATGTTGTTGCTGCTGATGACAGGATACACAGAGCTCATATG | 300 | |
| Qy | 878 | CATATAGCACTTTCTACATTTGGGTTTAAATATATATCAATSCAGATACCTTTTGGGAT | 937 | |
| Db | 301 | CATATAGCACTTTCTACATTTGGGTTTAAATATATATCAATSCAGATACCTTTTGGGAT | 360 | |
| Qy | 938 | TCACGCCAATCAGACAGTGAACACATGACGCTGAGGTCGCTTTCGATTCGTCGACAG | 997 | |
| Db | 361 | TCACGCCAATCAGACAGTGAACACATGACGCTGAGGTCGCTTTCGATTCGTCGACAG | 420 | |
| Qy | 998 | CTTATGCTCTTCTTGCACTATCTGAGAGCCGATTAACAAACCAAGAGTTCCAGACCTTT | 1057 | |
| Db | 421 | CTTACGCTTTTTCGAGTATCTGAGAGCCGTTTGAACAAACAGGAGTTCCAGACCTTT | 480 | |
| Qy | 1058 | TCCTTTTGGGTGATCACTAGCTGAGTGTGTTTCTTGTAGTGTATCTATTGACTT | 1117 | |
| Db | 481 | TCCTTTTGGGTGATCACTAGCTGAGGCGCTGTTTCTTGTAGTGTATCTATTGACTT | 540 | |
| Qy | 1118 | ATACAGGTTACATTCGACCATGAGTGGCGAGTTTATTCATTGTGGGATCTGGGTATG | 1177 | |
| Db | 541 | ACACAGGTTATATTCACCATGAGTGGCGAGTTTATTCATTGTGGGATCTGGGTATG | 600 | |
| Qy | 1178 | CAAAATACATTCACATTTATTCATGCTGAGTGTGAGTGTCACTACGACTTGGTGT | 1237 | |
| Db | 601 | CAAAATACATTCACATTTATTCATGCTGAGTGTGAGTGTCACTACGACTTGGTGT | 660 | |
| Qy | 1238 | CTTCTCTCTTTGATCTACATATTTCTGTATGTACCTTCCAGCAGGCTTTGGTTCTGCA | 1297 | |
| Db | 661 | CTTCTCTCTTTGATCTACATATTTCTGTATGTACCTTCCAGCAGGCTTTGGTTCTGCA | 720 | |
| Qy | 1298 | TCAAAATATCAAGATGAAGATTTTGTGCTCTATATGCAATCAGTGTGCTACT | 1357 | |
| Db | 721 | TCAAAATATCAAGATGAAGATTTTGTGCTCTATATGCAATCAGTGTGCTACT | 780 | |
| Qy | 1358 | TTGCTGGAGTGTGGTGGTGTGATTTGACTTTGACTCCAGTCCGTTGCTGCTG | 1417 | |
| Db | 781 | TTGCTGGAGTGTGGTGGTGTGATTTGACTTTGACTCCGTTGCTGCTG | 840 | |
| Qy | 1418 | CAATTGCTTTT | 1429 | |

| | | | | |
|-----------------------|---|--|-------|---------------------------------|
| Db | 841 | GCATCGCCTTCT | 852 | |
| RESULT 9 | | | | |
| BG698273 | | | | |
| LOCUS | | | | |
| DEFINITION | 602660188F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4803332 5', | 735 bp | mRNA | linear EST 07-MAY-2001 |
| ACCESSION | BG698273 | | | |
| VERSION | EST. | | | |
| KEYWORDS | EST. | | | |
| SOURCE | EST. | | | |
| ORGANISM | Homo sapiens (human) | | | |
| REFERENCE | 1 (bases 1 to 735) | | | |
| AUTHORS | NIH-WGC | http://mgc.ncl.nih.gov/ | | |
| TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) | | | |
| JOURNAL | Unpublished (1999) | | | |
| COMMENT | Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: James Cleaver, M.D. cDNA Library Preparation: Life Technologies, Inc. Sequencing by: Incyte Genomics, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10698 row: p column: 21 High quality sequence start: 2 High quality sequence stop: 734. | | | |
| FEATURES | 1..735 | | | |
| source | /organism="Homo sapiens" | | | |
| | /mol_type="mRNA" | | | |
| | /db_xref="taxon:9606" | | | |
| | /clone="IMAGE:4803332" | | | |
| | /lab_host="DH10B (T1 phage-resistant)" | | | |
| | /clone_lib="NCI CGAP Skn3" | | | |
| | /note="Organ: skin; Vector: pCMV-SPORT6; Site: i: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI CGAP Library." | | | |
| ORIGIN | | | | |
| Query Match | 29.3%; | Score 727; | DB 4; | Length 735; |
| Best Local Similarity | 99.3%; | Pred. No. 4e-182; | | |
| Mismatches | 730; | Conservative | 0; | Mismatches 5; Indels 0; Gaps 0; |
| Qy | 1434 | TGTTTTGACACTATTTCGGGGATGACATGAAGAGGAAATCCACCTGTGGAGCAG | 1493 | |
| Db | 1 | TGTTTTGACACTATTTCGGGGATGACATGAAGAGGAAATCCACCTGTGGAGCAG | 60 | |
| Qy | 1494 | CAGTGATGAGGATCAAAAGAAACCAAGAAATTTGTATGATAGGAGGTAAGTGAG | 1553 | |
| Db | 61 | CAGTGATGAGGATCAAAAGAAACCAAGAAATTTGTATGATAGGAGGTAAGTGAG | 120 | |
| Qy | 1554 | GAAACATGCACTGAACAGGAAACCTGAAGAGGATAGCCCTATATAAAGCAT | 1613 | |
| Db | 121 | GAAACATGCACTGAACAGGAAACCTGAAGAGGATAGCCCTATATAAAGCAT | 180 | |
| Qy | 1614 | TGTCACCATGTTGATGCTGATGCTATTGATGATGTTTCTGTCCACTGTACCTGGTCA | 1673 | |
| Db | 181 | TGTCACCATGTTGATGCTGATGCTATTGATGATGTTTCTGTCCACTGTACCTGGTCA | 240 | |
| Qy | 1674 | AAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCTGCTCATACATCATGATGACCCAG | 1733 | |
| Db | 241 | AAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCTGCTCATACATCATGATGACCCAG | 300 | |
| Qy | 1734 | GAATATCTTAGATGATTTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACA | 1793 | |
| Db | 301 | GAATATCTTAGATGATTTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACA | 360 | |

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 950)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 cDNA Library Preparation: Invitrogen Corp
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: Agencourt Bioscience Corporation
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM14085 row: b column: 03
 High quality sequence stop: 680.
 Location/Qualifiers
 1..950
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6512642"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_134"
 /note="Vector: pCMV-SPORT6.1; Site 1: EcorV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
 this is a NIH_MGC Library."

FEATURES
 source
 1..950
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6512642"
 /tissue_type="undifferentiated limb"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_134"
 /note="Vector: pCMV-SPORT6.1; Site 1: EcorV; Site 2: NotI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
 this is a NIH_MGC Library."

ORIGIN
 Query Match 30.9%; Score 767.6; DB 5; Length 950;
 Best Local Similarity 89.2%; Pred. No. 6.4e-193;
 Matches 838; Conservative 0; Mismatches 100; Indels 1; Gaps 1;
 643 TCAGTAGCTGGATCCTTTGATAATGAAGGCATTCCTATTTTGACATTCAGTTCAATAC 702
 Db 1 TCAGTGGCGGATCCTTTGATAATGAAGGCATTCCTATTTTGCGCTCAGTTCACTTAC 60
 703 TATTATGGGTAATCTGTAAACCTGGTCACTTTTGGGCAATGCTCTCTCTTA 762
 Db 61 TACTATGGGTAATCTGTGAAGCCGGTCTGTGTTCTGGCAATGCTCTCTCTG 120
 763 TCTATTTCTATATGCTCTCTGTTGGGCTGTTATGTTATTCATCAATCTTATCCA 822
 Db 121 TCATATTTCTATGCTCTCTGTTGGGCTGTTATGTTATTCATCACTCATCCT 180
 823 CTGCATGATTTGTTGTTTACTGATGAGAGATACAGCAAGAGTCTACATAT 882
 Db 181 CTCCATGTTGTTGTTGTTGCTGATGAGAGGATACAGCAAGAGTCTACATAT 240
 883 AGCATTCTTCAATGCTGGTTTATATTAATTCATGAGATACCTTTTGGGATCCAG 942
 Db 241 AGCATTCTTCAATGCTGGTTTATATTAATTCATGAGATACCTTTTGGGATCCAG 300
 943 CCATCAGACAGTGAACATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1002
 Db 301 CCATCAGACAGTGAACATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 1003 GC-TTCTTTCAGTATCTGAGACCCGATTAACAAAACAGAGTTCAGACCCCTTTCT 1061
 Db 361 GCTTTTTCAGTATCTGAGACCCGATTAACAAAACAGAGTTCAGACCCCTTTCT 420
 1062 TTGGGCTGATCCTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
 Db 421 TTGGGCTGATCCTAGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 1122 AGTTATATTCACCATGAGTGGCAGGCTTTTATTCATTTGGGATACCTGGTATGCAA 1181
 Db 481 AGTTATATTCACCATGAGTGGCAGGCTTTTATTCATTTGGGATACCTGGTATGCAA 540

QY 1182 AATACACATTCAATTAATTCATCAGTCTGTCAGCATCAACCTACGACTTGGGTGCTTT 1241
 Db 541 AATACACATTCAATTAATTCATCAGTCTGTCAGCATCAACCTACGACTTGGGTGCTTT 600
 QY 1242 CTTCTTTGATCTACATATCTTGTATGTATGATCTTCCAGCAGCCTTTGGTTCGATCAA 1301
 Db 501 CTTCTTTGATCTACATATCTTGTATGTATGATCTTCCAGCAGCCTTTGGTTCGATCAA 660
 QY 1302 AATATCAACGATGAAAGAGTATTTTGTCTTATATCAATCAATGCTGCTTACTTTC 1361
 Db 661 AATATCAACGATGAAAGAGTATTTTGTCTTATATCAATCAATGCTGCTTACTTTC 720
 QY 1362 TGAGGTGATGTCGACATGCTTCACTTGTACTTCACTGCTGCTGCTGCTGCTGCTGCT 1421
 Db 721 CNGAGTGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 QY 1422 TGCTTTTCAATGTTTTTGTAGTCACTATTTTGGGGATGACATGAAAGGGAATTCAC 1481
 Db 781 CGCTTTCTCAATGTTTTTGTAGTCACTATTTTGGGGATGACATGAAAGGGAATTCAC 840
 QY 1482 TGTGGAGGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1541
 Db 841 TGTGGAGGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 1542 AGTTAAGTGAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1580
 Db 901 AGTTAAGTGAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 939
 RESULT 8
 CN458732 852 bp mRNA linear EST 21-APR-2004
 LOCUS UI-M-HB0-cov-1-02-0-UI-r1 NIH_BMAP_HB0 Mus musculus cDNA clone
 DEFINITION IMAGE:30649513 5', mRNA sequence.
 ACCESSION CN458732
 VERSION CN458732.1 GI:46464458
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 852)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/mousefl.html
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 Location/Qualifiers
 1..852
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30649513"
 /tissue_type="whole eye"
 /dev_stage="embryo 12.5, 13.5, 14.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_HB0"
 /notes="Organ: Eye; Vector: pyx- Asc; Site 1: Ecor I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dt

QY 2111 TAGACAAAGCGAGATCCCTACTTTGTTGAATGCTTATGTTATATAAATGTCATACATCA 2170
 |||||
 Db 481 TAGACAAAGCGAGATCCCTACTTTGTTGAATGCTTATGTTATATAAATGTCATACATCA 540
 |||||
 QY 2171 GATTGGAGAAATGAGCTGCTGATTTTCGTACACCCCGAGGTTTGGACCGAACAACGTAATG 2230
 |||||
 Db 541 GATTGGAGAAATGAGCTGCTGATTTTCGTACACCCCGAGGTTTGGACCGAACAACGTAATG 600
 |||||
 QY 2231 CTGAGATTGGAAATTAAGACATTAATTAATCAAAATTTGGAGAAAGCTTTTACATCAGAAC 2290
 |||||
 Db 601 CTGAGATTGGAAATTAAGACATTAATTAATCAAAATTTGGAGAAAGCTTTTACATCAGAAC 660
 |||||
 QY 2291 ACTGGCTTGTAGGATATATAAAGTAAAGCACTGTATAACAGGAGAGACATTAGATCACA 2350
 |||||
 Db 661 ACTGGCTTGTAGGATATATAAAGTAAAGCACTGTATAACAGGAGAGACATTAGATCACA 720
 |||||
 QY 2351 AACCTCGAGTCACCAACATTTTCCAAACAGAGATTTTGTCAAAGAGACTTACCAAAA 2410
 |||||
 Db 721 AACCTCGAGTCACCAACATTTTCCAAACAGAGATTTTGTCAAAGAGACTTACCAAAA 780
 |||||
 QY 2411 GGAAGCGTGGCTACATTAATAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAGA 2470
 |||||
 Db 781 GGAAGCGTGGCTACATTAATAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAGA 840
 |||||
 QY 2471 AGACTGTATAA 2481
 |||||
 Db 841 AGACTGTATAA 851
 |||||

RESULT 6

EX363211 1088 bp mRNA linear EST 08-APR-2004
 LOCUS EX363211 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 DEFINITION CDNA clone CS0DK010Y001 5-PRIME, mRNA sequence.

ACCESSION EX363211

VERSION EX363211.2 GI:46291519

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1088)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On May 5, 2003 this sequence version replaced gi:30372696.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 2459.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?c=CS0DK010AH01QP1&c=2459.f.

FEATURES

source

1..1088

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK010Y001"

/cell_type="HELA CELLS"

/clone_lib="HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized."

ORIGIN

Query Match 32.6%; Score 809.8; DB 5; Length 1088;
 Best Local Similarity 96.2%; Pred. No. 3.7e-204;
 Matches 821; Conservative 24; Mismatches 6; Indels 2; Gaps 2;
 QY 1631 TGATGCTATTGATGTTGTTGCTGCTCCACTGTACCTGGGTCCACAGCAATGCTTACTCTA 1690
 |||||
 Db 1 TGATGCTATTGATGTTGTTGCTGCTCCACTGTACCTGGGTCCACAGCAATGCTTACTCTA 60
 |||||
 QY 1691 GTCCTAAGTGTAGTCTGCTGCTCATACATCATCATGTCGCCACGAGGAATATCTTAGATGATT 1750
 |||||
 Db 61 GTCCTAAGTGTAGTCTGCTGCTCATACATCATCATGTCGCCACGAGGAATATCTTAGATGATT 120
 |||||
 QY 1751 TTAGAGAAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAAATGCAATGCAAGTAAATGCTTT 1810
 |||||
 Db 121 TTAGAGAAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAAATGCAAGTAAATGCTTT 180
 |||||
 QY 1811 GGTGGGATTTGGCTATCAGATAGCTGGAATGGCTTAATAGAACTACCTTGGTGGATAATA 1870
 |||||
 Db 181 GGTGGGATTTGGCTATCAGATAGCTGGAATGGCTTAATAGAACTACCTTGGTGGATAATA 240
 |||||
 QY 1871 ACACCTGGAAATACAGCCACATAGCACCTGGTGGGAAAAGCTTATGCTTCTTAATGAAACAG 1930
 |||||
 Db 241 ACACCTGGAAATACAGCCACATAGCACCTGGTGGGAAAAGCTTATGCTTCTTAATGAAACAG 300
 |||||
 QY 1931 CAGCCTTAATAATCATGAGGACTCTAGATGATAGTATGTTTGGTATTTTGGAGGGG 1990
 |||||
 Db 301 CAGCCTTAATAATCATGAGGACTCTAGATGATAGTATGTTTGGTATTTTGGAGGGG 360
 |||||
 QY 1991 TTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGAGGTTTAGGATAGCTGAAG 2050
 |||||
 Db 361 TTATTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGAGGTTTAGGATAGCTGAAG 420
 |||||
 QY 2051 GAGAAATCCCAAGACATTTGGGAAAGTACTATTTACCCACAGGAGGAAATTCGGTG 2110
 |||||
 Db 421 GAGAAATCCCAAGACATTTGGGAAAGTACTATTTACCCACAGGAGGAAATTCGGTG 480
 |||||
 QY 2111 TAGACAAAGCGAGATCCCTACTTTGTTGAATGCTTATGTTATATAAATGTCATACATCA 2170
 |||||
 Db 481 TAGACAAAGCGAGATCCCTACTTTGTTGAATGCTTATGTTATATAAATGTCATACATCA 540
 |||||
 QY 2171 GATTGGAGAAATGAGCTGCTGATTTTCGTACACCCCGAGGTTTGGACCGAACAACGTAATG 2230
 |||||
 Db 541 GATTGGAGAAATGAGCTGCTGATTTTCGTACACCCCGAGGTTTGGACCGAACAACGTAATG 600
 |||||
 QY 2231 CTGAGATTGGAAATTAAGACATTAATTAATCAAAATTTGGAGAAAGCTTTTACATCAGAA 2289
 |||||
 Db 601 CTGAGATTGGAAATTAAGACATTAATTAATCAAAATTTGGAGAAAGCTTTTACATCAGAA 660
 |||||
 QY 2290 CACTGGCTTGTAGGATATATAAAGTAAAGCACTG-ATACAGGAGAGACATTAGATCA 2348
 |||||
 Db 661 CACTGGCTTGTAGGATATATAAAGTAAAGCACTG-ATACAGGAGAGACATTAGATCA 720
 |||||
 QY 2349 CAAACCTCGAGTCACCAACATTTTCCAAACAGAGATTTTGTCAAAGAGACTTACCAA 2408
 |||||
 Db 721 CAAACCTCGAGTCACCAACATTTTCCAAACAGAGATTTTGTCAAAGAGACTTACCAA 780
 |||||
 QY 2409 AAGGAGAGCTGCTGCTACATTAATAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAA 2468
 |||||
 Db 781 AAGGAGAGCTGCTGCTACATTAATAAATTAAGCTGGTTTAAAGAAAGGCAAGAAATATCTAA 840
 |||||
 QY 2469 GAAGACTGTTAA 2481
 |||||
 Db 841 GAAGACTGTTAA 853
 |||||

RESULT 7

BU515954

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

BU515954 950 bp mRNA linear EST 12-SEP-2002

AGENCOURT_10120433 NIH_MGC_134 Mus musculus cDNA clone

IMAGE:6512642 5', mRNA sequence.

BU515954

BU515954

GI:22823480

EST.

QY 1394 CTCAGTCGTGTATGCTGTCTGCAATTCGCTTTTCAAAATGTTTTTGAGCACTATTTCG 1453
 Db 1053 ----- 1052
 QY 1454 GGGATGACATGAAGGAAATCCACTGTGGAGGACAGCAGTGATGAGATGACAAA 1513
 Db 1053 ----- 1052
 QY 1514 GAAACCAAGAAATTTGTATGATGAAGCAGGTAAAGTGAGAAACATGCAACTGACACAG 1573
 Db 1053 -----GGCAGGAAAGTGAGAAAGCATGTGACGGACAG 1087
 QY 1574 AAAAACTGAAGGGATTAGCCCTTAATATAAAAGCATGTGACCATGTGATGCTGA 1633
 Db 1088 AGAAGCCGGAGGAGGCTTGGCCCAACATCAAGAGCATCGTGACCATGCTCATGCTCA 1147
 QY 1634 TGCTATTGATGATGTTGCTCTCCACTGTACTTGGGTCAACAGCAATGCTACTCTAGTC 1693
 Db 1148 TGCTCTCATGATGTTGCTGTGCACTGCACTGGGTCAACAGCAAGGCTACTCCAGCC 1207
 QY 1694 CAAGTGATGCTGGCTCTACATCATGATGGCACCAGGAATATCTTAGATGATTTA 1753
 Db 1208 CAAGTGATGCTGGCTCTACATCATGATGGCACCAGGAATATCTTAGATGATTTA 1267
 QY 1754 GAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTTGGT 1813
 Db 1268 GAGAGCACTACTTTTGGTTGAGACAAACACGGATGAACACGCGCACTGCTGCTGT 1327
 QY 1814 GGGATTATGCTATCAGATGCTGGATGCTATGATGATGATGATGATGATGATGATGAT 1873
 Db 1328 GGGATATGCTATCAGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGAT 1387
 QY 1874 CTTGGAATPAACAGCCACATAGCACTGTTGGGAAAGCATGCTTCTTAATGAACAGCAG 1933
 Db 1388 CTTGGAACACACAGCCACATGCTGTTAGGAAAGCATGCTTCTTAATGAACAGCAG 1447
 QY 1934 CCTATTAATCATGAGGCTCTAGATGATGATGATGATGATGATGATGATGATGATGAT 1993
 Db 1448 CCTATTAATCATGAGGCTCTAGATGATGATGATGATGATGATGATGATGATGATGAT 1507
 QY 1994 TTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGATGTTAGGATAGCTGAAGGAG 2053
 Db 1508 TTGGCTATTCTGGTGATGATATCAACAAATTTCTCTGGATGTTAGGATAGCTGAAGGAG 1567
 QY 2054 AACATCCCAAGACATTCGGGAAAGTCACTATTATCCCAAGGAGGATTCCTGGTAG 2113
 Db 1568 AGCACCACCAAGACATTCGGGAAAGTCACTATTATCCCAAGGAGGATTCCTGGTAG 1627
 QY 2114 ACAAAGCAGGATCCCTACTTTGTTGAATTCCTTATGATATAAAATGTCATACAGAT 2173
 Db 1628 ACAGAGCCGGTCTCTACTCTGTTAAATTCCTTATGATATAAAATGTCATACAGAT 1687
 QY 2174 TTGGAGAAATCAGCTGATTTTCGTACACCCCGAGTTTGGACCAACAGCTTAATGCTG 2233
 Db 1688 TTGGAGAAATCAGCTGATTTTCGTACACCCCGAGTTTGGACCAACAGCTTAATGCTG 1747
 QY 2234 AGATTGGAATTAAGACATTAATTAATTAAGATTCGGAAGCCTTTTACATCAGACACT 2293
 Db 1748 AGATTGGAATTAAGACATTAATTAATTAAGATTCGGAAGCCTTTTACATCAGACACT 1807
 QY 2294 GCTTGTAGGATATATAAGTAAAGCAGCTGATTAACAGGAGAGCATTAGATCAACAC 2353
 Db 1808 GCTTGTAGGATATATAAGTAAAGCAGCTGATTAACAGGAGAGCATTAGATCAACAC 1867
 QY 2354 CTCGAGTCCCAACATTTTCCCAACAGAGGATTTTGTCAAGAGAG 2400
 Db 1868 CTCGAGTCCCAACATTTTCCCAACAGAGGATTTTGTCAAGAGAG 1914

RESULT 5
 CR610915
 LOCUS

DEFINITION full-length cDNA clone CS0DK010Y001 of HeLa cells Cot 25-normalized
 of Homo sapiens (human).
 ACCESSION CR610915
 VERSION CR610915.1 GI:50491722
 KEYWORDS HTC; CDS; CDS; CDS
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2212)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 REMARK Contact : feng liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue
 REFERENCE 2 (bases 1 to 2212)
 Genoscope.
 Direct Submission
 Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 Location/Qualifiers
 1..2212
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK010Y001"
 /issue_type="HeLa cells Cot 25-normalized"
 /plasmid="pCMVSPORT_6"
 ORIGIN
 Query Match 34.3%; Score 851; DB 3; Length 2212;
 Best Local Similarity 100.0%; Fred. No. 4.7e-215; Indels 0; Gaps 0;
 Matches 851; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1631 TGATGCTATTGATGATGATTTGCTGCTCCACTGACCTGGGTGTCACAGCAATCTACTCTA 1690
 Db 1 TGATGCTATTGATGATGATTTGCTGCTCCACTGACCTGGGTGTCACAGCAATCTACTCTA 60
 QY 1691 GTCCAGCTGACCTGCTGCTCATCAATCATGATGGCAGCAGCAATCTTAGATGATT 1750
 Db 61 GTCCAGCTGATGATGCTGCTCATCAATCATGATGGCAGCAGCAATCTTAGATGATT 120
 QY 1751 TTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTT 1810
 Db 121 TTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTT 180
 QY 1811 GGTGGGATTTGCTATCAGATGATGCTGGATGGCTTAATAGCACTACGTTGGTGGTAATA 1870
 Db 181 GGTGGGATTTGCTATCAGATGATGCTGGATGGCTTAATAGCACTACGTTGGTGGTAATA 240
 QY 1871 ACACCTGGAATAACAGCCACATAGCACTGCTGGTGGGAAAAGCTATGCTTCTTAATGAACAG 1930
 Db 241 ACACCTGGAATAACAGCCACATAGCACTGCTGGTGGGAAAAGCTATGCTTCTTAATGAACAG 300
 QY 1931 CAGCCTATAAATCATGAGGACTCTAGATGATGATGATGATGATGATGATGATGATGATGAT 1990
 Db 301 CAGCCTATAAATCATGAGGACTCTAGATGATGATGATGATGATGATGATGATGATGATGAT 360
 QY 1991 TTATGGCTATTCTGCTGATGATATCAACAAATTTCTCTGAGTGGTGGTGGTGGTGGTGG 2050
 Db 361 TTATGGCTATTCTGCTGATGATATCAACAAATTTCTCTGAGTGGTGGTGGTGGTGGTGG 420
 QY 2051 GAGACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGAGCAATTCCTG 2110
 Db 421 GAGACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGAGCAATTCCTG 480

/note="Vector: PKMV-SPORT6"

ORIGIN

Query Match 66.1%; Score 1640.6; DB 3; Length 4513;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1911; Conservative 0; Mismatches 404; Indels 8; Gaps 2;

```
QY 159 GGCCTCGAAGACATCATCCAGAGTTCGACCCGCTGCTGCGAGCGCTGCTGCGAGTCGCTCTCTC 218
DB 414 GGGCCATGGAGACATGTCGGCGGCTTGTGCGACGCGCGGCTGCGAGTCTCTGCTCTC 473
QY 219 CTTACACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 278
DB 474 CTTACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 533
QY 279 CGCTTCGAAACATCATCCAGAGTTCGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338
DB 534 CGCTTCGAGACATCATCCAGAGTTCGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
QY 339 TCTTGCACTCTCATGGTCTATGAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 398
DB 594 CTTAGCATCCCATGCTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 653
QY 399 ACTAGGAAGATAGTAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 458
DB 654 ACTGGGAGATAGTAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 713
QY 459 TCATTGGATTTTAAATCACTTGAACATTAACCTGTTCAATAGAGAGATAGTGTCTCTCTCTCT 518
DB 714 CCATTGGATTTTAAATCACTTGAACATTAACCTGTTCAATAGAGAGATAGTGTCTCTCTCTCT 773
QY 519 TGCACCACTTTTAGCGGCTTACATCTATCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 578
DB 774 AGCACCAGATTTTAGTGGCTTACATCAATCTCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCT 833
QY 579 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
DB 834 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893
QY 639 TCGGTCACTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
DB 894 CAGATCAGTGGATCATTTGACATTAAGAGGATGCTATCTTGCATCTGAGTTAC 953
QY 699 ATACTATTATCGGTAAATCTGTAACAACTGGGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 758
DB 954 ATATTATTGCGGTAAATCTGTAACAACTGGGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1013
QY 759 CTTATCTATTCTATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 818
DB 1014 TCTGCTTATTTTATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1073
QY 819 TCCACTGCACTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 878
DB 1074 TCCACTGCACTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1133
QY 879 ATATAGCACTTCTCATAGTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 938
DB 1134 TTACAGCACTTTTATAGTGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1193
QY 939 CCAGCCAAATCAGAACTGAAACATGCGAGCTGCGAGTCTCTTTCATTTGCTGCAAGC 998
DB 1194 TCAGCCAAATCAGAACTGAAACATGCGAGCTGCGAGTCTCTTTCATTTGCTGCAAGC 1253
QY 999 TTATGCTTTCTTCTGAGTATCTGAGAGACCGATTAACAAACAGAGATTCAGACCTTTT 1058
DB 1254 CTATGCTTTCTTCTGAGTATCTGAGAGACCGATTAACAAACAGAGATTCAGACCTTTT 1313
QY 1059 CTTTTTGGGTGATCACTAGTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
DB 1314 TTTCTCTGCTGATCGTTGGGAGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1373
QY 1119 TACAGTTCATGACCACTGAGTGGCAGGTTTATCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1178
```

```
DB 1374 TACAGGCTACATTTGCTCCATGGAGTGGAGATTTTACTCATTCTGGGACACAGGATATGC 1433
QY 1179 AAAAAATACATTTCCAAATTTTGCATTCAGTGTCTGAGCATCAACTAGCATTTGGGTGTC 1238
DB 1434 GAAAAATCCATATCCCATCATTTGCTGTGTCTGAGCATCAACCAACCAATGGGTGTC 1493
QY 1239 TTTCTCTTTGATCTACATATTTCTGTATGTACCTTTCCAGCAGGCTTTGGTCTGTCAT 1298
DB 1494 CTTCTTTTGTGATCTGATATATTAGTATGTACATTTCTCGCAGGACTTTGGTCTGTCAT 1553
QY 1299 CAAAAATATCAAGATGAAAGATTTTGTGCTCTATATGCAATCAGTCTGCTCTACTT 1358
DB 1554 AAAAAATATCAAGATGAAAGATTTTGTGCTCTGATGCAATTAGTCCGTTACTT 1613
QY 1359 TGTGAGTGTGATGCTGAGTGTGATGTTGACTTTCAGTCCAGTCTGCTGCTGCTGCTGCTGCT 1418
DB 1614 TGTGCGGTAATGGTTCGTTGATGCTGACCTTGACTCTCTGCTGCTGCTGCTGCTGCTGCT 1673
QY 1419 AATTGCTTTTCAAAATTTTGTGACACTTATTTGGGGATGACATGAAAGAGAAATCC 1478
DB 1674 CATGCGCTTCTCCAAATGTTTGTGAGCATTTCTGGGTGATGATGAAAGCGGAAACCC 1733
QY 1479 ACTGTGGAGGACAGCAGTGTGAGGATGACAAAAAGAAACCAAGGAAATTTGTATGATAA 1538
DB 1734 CCCAGTGGAAAGACAGCAGTGAACAAAG-----AGCGAAATTTCTGGTACACTCTATGATAA 1787
QY 1539 GGCAGGTAAAGTGCAGAAACATGCAACTGAACAGGAAABAACTGAAGAGGATTTAGGCC 1598
DB 1788 GGCAGGAAAGTGCAGAAACATGTCAGGAGGAGAAAGCAGAGGAGTCTTGGGCC 1847
QY 1599 TAAATATAAAGCAATGTCACTATGCTGATGCTGATGCTATGATGATGATGATGATGATGATGAT 1658
DB 1848 TAAATATAAAGCAATGTCACTATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1907
QY 1659 CTGTACTGCTGCTCACAAGCAATGCTCTACTCTAGTCCAAAGTGTAGTCTGCTGCTGCTCATACAA 1718
DB 1907 CTGTACTGCTGCTCACAAGCAATGCTCTACTCTAGTCCAAAGTGTAGTCTGCTGCTGCTCATACAA 1967
QY 1719 TCAATGAGGACCAAGCAATGCTCTAGTATGATTTTGAAGAGCTTACTTTTGGTGAAGCA 1778
DB 1967 TCAATGAGGACCAAGCAATGCTCTAGTATGATTTTGAAGAGCTTACTTTTGGTGAAGCA 2027
QY 1779 AATACAGATGAACATGTCAGAGTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGG 1838
DB 2027 GAATACAGATGAACATGTCAGAGTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGG 2087
QY 1839 AATGCTTAATAGAACTAGTGTGGATTAATTAACACCTGGGAATTAACAGCCACATAGCACT 1898
DB 2087 AATGCTTAATAGAACTAGTGTGGATTAATTAACACCTGGGAATTAACAGCCACATAGCACT 2147
QY 1899 GGTGGGAAAAAGCTATGCTCTTAAATGAAACAGCAGCTTATAAAATCATGAGGACTCTAGA 1958
DB 2148 GGTGGGAAAAAGCTATGCTCTTAAATGAAACAGCAGCTTATAAAATCATGAGGACTCTAGA 2207
QY 1959 TGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGCTGATGATATCAA 2018
DB 2207 TGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGCTGATGATATCAA 2267
QY 2019 CAAATTTCTCTGGATGTTTGGATAGTCTCAAGAGGAGAAATCCCAAGAGCAATTTCCGGAAAG 2078
DB 2267 CAAATTTCTCTGGATGTTTGGATAGTCTCAAGAGGAGAAATCCCAAGAGCAATTTCCGGAAAG 2327
QY 2079 TGACTATTTTACCCACAGGAGAAATTCGCTGTAGCAAAAGCAGGATCCCTACTTTGTT 2138
DB 2327 TGACTATTTTACCCACAGGAGAAATTCGCTGTAGCAAAAGCAGGATCCCTACTTTGTT 2385
QY 2139 GAATTTGCTTTATGATAAAATGTCTACTATCAAGATTTGGAGAAATSCAGCTGGATTTTCG 2198
DB 2385 GAATTTGCTTTATGATAAAATGTCTACTATCAAGATTTGGAGAAATSCAGCTGGATTTTCG 2445
QY 2199 TACACCCCGAGGTTTTCAGCGACAGCTAATCTCGAGTTTGGTGAATGAGCATTAATTT 2258
DB 2445 TACACCCCGAGGTTTTCAGCGACAGCTAATCTCGAGTTTGGTGAATGAGCATTAATTT 2505
```

Db 1046 GAGAGACCGGTTGACAAACAGGAGTCCAGACCCCTTTCTTTCTTTGGGTGTCTCACTAGC 1105
Qy 1080 TGCAGGTGTGTGTTCTTCTAGTGTCACTATTTTCACTTATACAGGTGTACATTCACCATG 1139
Db 1106 TGCAGGCGTGTGTCTTCTAGTGTCACTATTTTCACTTATACAGGTGTATATTCACCATG 1165
Qy 1140 GAGTGCAGGTTTATTCATTTGTGGATACCTGGGTATGCAAAATACACATTCCTCAATAT 1199
Db 1166 GAGTGCAGGTTTATTCATTTGTGGATACCTGGGTATGCAAAATACACATTCCTCAATAT 1225
Qy 1200 TGCATCAGTGTCTGAGCATCACTTACGACTTGGGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1259
Db 1226 TGCATCAGTGTCTGAGCATCACTTACGACTTGGGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1285
Qy 1260 TCTTGTATGCTCTTCCAGCAGGCTTTTGTCTTCTGATCAAAATATCAACGATGAAG 1319
Db 1286 TCTTGTATGCTCTTCCAGCAGGCTTTTGTCTTCTGATCAAAATATCAACGATGAAG 1345
Qy 1320 AGTATTTGTCTCTATATGCAATCAGTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1379
Db 1346 AGTATTTGTCTCTATATGCAATCAGTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1405
Qy 1380 GATGTGTACTTGTACTCCAGCTGTGTATGCTCTTCTGCAATTCCTTTTCAATGTTTT 1439
Db 1406 GATGTGTACTTGTACTCCAGCTGTGTATGCTCTTCTGCAATTCCTTTTCAATGTTTT 1465
Qy 1440 TGAGCAGTATTTGGGGGATGATCAATGAAGGAAATCCACCTGTGGAGCAGCAGTGA 1499
Db 1466 TGAGCAGTATTTGGGGGATGATCAATGAAGGAAATCCACCTGTGGAGCAGCAGTGA 1525
Qy 1500 TGAGGATGCAAAAGAAACCAAGGAAATTTGATGATGAGCAGGTAAGTGAAGGAAAC 1559
Db 1526 TGAGGATGCAAAAGAAACCAAGGAAATTTGATGATGAGCAGGTAAGTGAAGGAAAC 1585
Qy 1560 TCCACTGACAGGAAAGAACTGAGAGGAAATGAGGCTTAAATATATAAAGCATTCTAC 1619
Db 1586 TGTGACAGCAGCAGGAAACCTGAGAGGCTTAAATATATAAAGCATTCTAC 1645
Qy 1620 CATGTGTATGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1679
Db 1646 CATGTGTATGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1705
Qy 1680 TGCCTACTCTAGTCCAAAGTGTAGTCTGCGCTCATACAACTGATGAGCAGCAGGAAAT 1739
Db 1706 CCGCTACTCTCAAGTGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1765
Qy 1740 CTTAGATGATTTAGAGAGCTTACTTTTGTGCTTAAAGCAAAATACAGATGAACGACG 1799
Db 1766 ATTAGATGATTTAGAGAGCTTACTTTTGTGCTTAAAGCAAAATACAGATGAACGACG 1825
Qy 1800 AGTATGCTTGTGCTGGGATTTGGTATCAGATAGTGGATGGCTAATAGAACTACGTT 1859
Db 1826 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1885
Qy 1860 GTTGAATATACACTGTGAATACAGCCATAGCATGCTGGTGGGAAAGCTATGCTTC 1919
Db 1886 GTTGAATATACACTGTGAATACAGCCATAGCATGCTGGTGGGAAAGCTATGCTTC 1945
Qy 1920 TAATGAACAGCAGGCTTAAATCATGAGGACTTATAGTGTAGATGATGATGATGATGAT 1979
Db 1946 CAATGAACGCGCTTAAATCATGAGGCTTCTTGTGCTGATGATGATGATGATGATGAT 2005
Qy 1980 TTTTGGAGGGTATGCTTATCTGTTGATGATATCAAAATTTCTCTGATGTTAG 2039
Db 2006 TTTTGGAGGGTATGCTTATCTGTTGATGATATCAAAATTTCTCTGATGTTAG 2065
Qy 2040 GATAGCTGAAGGAGAACATCCCAAGACATTCGCGAAAGTGAATTTTACCCACAGGG 2099
Db 2066 GATAGCTGAAGGAGAACATCCCAAGACATTCGCGAAAGTGAATTTTACCCACAGGG 2125
Qy 2100 AGAATTCGCTGATGACAAAGAGGATCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2159
Db 2126 AGAATTCGCTGATGACAAAGAGGATCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2185

Qy 2160 GTCTACTACAGATTGGGAAATGACGCTGGATTTTCGTACACCCCGGTTTGGACCG 2219
Db 2186 GTCTACTACAGATTGGGAAATGACGCTGGATTTTCGTACACCCCGGTTTGGACCG 2245
Qy 2220 AACACSTAAATGCTGAGATTGGAATTAAGCAATTAAATTTCAAAATTTGGAAGAGCCTT 2279
Db 2246 AACACSTAAATGCTGAGATTGGAATTAAGCAATTAAATTTCAAAATTTGGAAGAGCCTT 2305
Qy 2280 TACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCAACCTGATTAACAGGGAGAC 2339
Db 2306 TACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCAACCTGATTAACAGGGAGAC 2365
Qy 2340 ATTAGATCAAAACCTCGAGTCACCAATTTTCCAAAACAGAAATTTTGAAGAA 2399
Db 2366 ACTAGGTCAAAACCTCGAGTCACCAATTTTCCAAAACAGAAATTTTGAAGAA 2425
Qy 2400 GACTACCAAAAGAGAGCGTGGCTACATTAATAAATTAAGCTGTTTAAAGAAAGCAAGAA 2459
Db 2426 GACTACCAAAAGAGAGCGTGGCTACATTAATAAATTAAGCTGTTTAAAGAAAGCAAGAA 2485
Qy 2460 AATATCTAAAGAGAGCTGTTTAA 2481
Db 2486 GACCTCTAAAGAGAGCTGTTTAA 2507

RESULT 3
BC044321

LOCUS BC044321 4513 bp mRNA linear HTC 23-JAN-2003
DEFINITION Xenopus laevis, clone IMAGE:4684177, mRNA.
ACCESSION BC044321
VERSION BC044321.1 GI:27882209
KEYWORDS HTC.

SOURCE

ORGANISM Xenopus laevis (African clawed frog)
Xenopus laevis
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
1 (bases 1 to 4513)
Klein, S. and Strausberg, R.
Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene
Collection (XGC), National Institute of Child Health and Human
Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
20892-7510, USA
NIH-MGC Project

REMARK

COMMENT Contact: XGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: anadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Series: IRAC plate: 94 Row: i Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, similarity but not identity to protein
This clone has the following problem: frame shifted.

FEATURES

source 1..4513
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4684177"
/tissue_type="Embryo, stage 31/32, Xenopus"
/clone_lib="NCHD_XGC_Emb4"
/lab_host="DH10B"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:24:44 ; Search time 7543.8 seconds
(without alignments)
11984.275 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcggccccc.....tatctaagaagacttttaa 2481

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hcc.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gss1.*

9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|------------|--------------------|
| 1 | 2078.6 | 83.8 | 2709 | 3 AK018758 | AK018758 Mus muscu |
| 2 | 2055.6 | 82.9 | 2669 | 3 AK012153 | AK012153 Mus muscu |
| 3 | 1640.6 | 66.1 | 4513 | 3 BC044321 | BC044321 Xenopus l |
| 4 | 981.8 | 39.6 | 2334 | 3 AY310157 | AY310157 Rattus no |
| 5 | 851 | 34.3 | 2212 | 3 CR610915 | CR610915 full-len |
| 6 | 809.8 | 32.6 | 1088 | 5 BX363211 | BX363211 BX363211 |
| 7 | 767.6 | 30.9 | 950 | 5 BU515934 | BU515934 AGENCOURT |
| 8 | 731.2 | 29.5 | 852 | 7 CN458732 | CN458732 UI-M-HO- |
| 9 | 727 | 29.3 | 735 | 4 BG698273 | BG698273 602660188 |
| 10 | 698.4 | 28.1 | 879 | 5 EQ947308 | EQ947308 AGENCOURT |
| 11 | 695.8 | 28.0 | 823 | 3 CF736482 | CF736482 UI-M-HO- |
| 12 | 692.6 | 27.9 | 778 | 4 BG573592 | BG573592 602594608 |
| 13 | 686 | 27.7 | 867 | 4 B1851585 | B1851585 603378796 |
| 14 | 685.2 | 27.6 | 861 | 7 CF995035 | CF995035 AGENCOURT |
| 15 | 674.6 | 27.2 | 798 | 7 CN537592 | CN537592 UI-M-HO- |
| 16 | 667.4 | 26.9 | 3802 | 3 AK030363 | AK030363 Mus muscu |
| 17 | 663.4 | 26.7 | 2656 | 3 BC028897 | BC028897 Mus muscu |
| 18 | 658 | 26.5 | 2730 | 3 AK031547 | AK031547 Mus muscu |
| 19 | 655.4 | 26.4 | 3643 | 3 AK077877 | AK077877 Mus muscu |
| 20 | 649.6 | 26.2 | 754 | 7 CK636437 | CK636437 UI-M-HO- |
| 21 | 648.8 | 26.2 | 814 | 7 CN531243 | CN531243 UI-M-HO- |
| 22 | 640.2 | 25.8 | 766 | 7 CK633751 | CK633751 UI-M-HO- |
| 23 | 634 | 25.6 | 877 | 6 CB208902 | CB208902 AGENCOURT |
| 24 | 633.4 | 25.5 | 757 | 4 B1869525 | B1869525 603392966 |

| | | | | | |
|----|-------|------|------|------------|--------------------|
| 25 | 623.6 | 25.1 | 738 | 7 CN527774 | CN527774 UI-M-HO- |
| 26 | 621 | 25.0 | 759 | 4 B1662821 | B1662821 603286252 |
| 27 | 620.4 | 25.0 | 2103 | 9 AY418286 | AY418286 Mus muscu |
| 28 | 618.6 | 24.9 | 691 | 5 BX669060 | BX669060 BX669060 |
| 29 | 618.6 | 24.9 | 2118 | 9 AY418284 | AY418284 Homo sapi |
| 30 | 617.6 | 24.9 | 957 | 5 BU554374 | BU554374 AGENCOURT |
| 31 | 616.4 | 24.8 | 776 | 4 BG620025 | BG620025 602618240 |
| 32 | 615 | 24.8 | 627 | 4 BG434322 | BG434322 602506307 |
| 33 | 614.8 | 24.8 | 825 | 6 CA318063 | CA318063 UI-M-FW- |
| 34 | 614 | 24.7 | 2727 | 3 AK087470 | AK087470 Mus muscu |
| 35 | 613.4 | 24.7 | 709 | 7 CF780700 | CF780700 AGENCOURT |
| 36 | 613 | 24.7 | 903 | 6 CD358477 | CD358477 AGENCOURT |
| 37 | 611.8 | 24.7 | 779 | 5 BM963786 | BM963786 UI-M-EQ- |
| 38 | 611.8 | 24.7 | 779 | 6 CD351656 | CD351656 UI-M-GIO- |
| 39 | 611.4 | 24.6 | 775 | 7 CK032376 | CK032376 AGENCOURT |
| 40 | 611 | 24.6 | 747 | 2 BE952197 | BE952197 UI-M-CD- |
| 41 | 611 | 24.6 | 2419 | 3 CNS0ADY2 | CNS0ADY2 Arabidops |
| 42 | 609.8 | 24.6 | 733 | 6 CA321357 | CA321357 UI-M-FW- |
| 43 | 609.6 | 24.6 | 689 | 2 BF970812 | BF970812 602273941 |
| 44 | 609.4 | 24.6 | 702 | 4 B1254237 | B1254237 602974710 |
| 45 | 609.4 | 24.6 | 724 | 7 CF737921 | CF737921 UI-M-HO- |

ALIGNMENTS

RESULT 1

AK018758

LOCUS

DEFINITION

AK018758

2709 bp

mRNA

linear

HTC 03-APR-2004

Mus musculus adult male liver cDNA, RIKEN full-length enriched

library, clone:1300006C19 product:hypothetical Oligosaccharyl

transferase (OTase) STT3 subunit containing protein, full insert

sequence.

AK018758

AK018758.2

GI:26384577

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;

QY 1503 GGATGACAAAAGAACCAAGGAAATTTGTATGATAGCGAGGTAAAGTGAGGAAAACATGC 1562
DB |||||
QY 1356 TGATGACCTTAGTCAGCAATCAAGAGCTCAACCAAAAATCAAGTTCAGATTTCCAAAT 1415
DB |||||
QY 1563 AACTGAACAGGAAAACCTGAAGAGGATAGGCCCTTAATATAAAAAGCATGTGCACCAT 1622
DB |||||
QY 1416 TGTGGATATTTGTCAAAAGTTTTGTTTACTGACATTTACATTTTACCTTTCTAC-- 1473
DB |||||
QY 1623 GTTGATGCTGATGCTATTGATGATGTTTGTGTCCTACTGCTACTGCTGGGTCAAGCAATGC 1682
DB |||||
QY 1474 -----TTGTTTACATTTGATTTGGGTAACTTGAATGCTGCAATGC 1508
DB |||||
QY 1683 CTACTCTAGTCCAAAGTGTAGCTGGCTCATACAAATCATGATGGCAGGCAATATCTT 1742
DB |||||
QY 1509 TTAATCATCAATCAGTTGTTTAGCATCCAGAAACCCAGATGGCTCAACATATCAT 1568
DB |||||
QY 1743 AGATGATTTAGAGAGCTTACTTTTGGCTTAGGCAAAATACAGATGAACATGCAGAT 1802
DB |||||
QY 1569 TGATGATATAGAGAGCTTACTTACTGGTTAAGAAATGAATACACCAAGAGATGCCAAAGT 1628
DB |||||
QY 1803 AATGCTTTGGTGGGATTTAGCTATACAGATAGCTTGAATGGCTAAATAGAACTACGTTGGT 1862
DB |||||
QY 1629 TATGCTGTGGGATTTATGTTATCAATCGGGGTATGGCTGATAGAACCCACACTGT 1688
DB |||||
QY 1863 GGATTAATACACCTGGATATACACCCACATAGCACTGGTGGGAAAAGCTATGCTCTAA 1922
DB |||||
QY 1689 TGATTAACAAATACATGGAATAACACATATGCGACTGTTGGTAAGCAATGCTTTCCOC 1748
DB |||||
QY 1923 TGAAACAGCAGCCTATAAAATCATGAGGACTCTAGATGTAGATATGTTTGGTTATTTT 1982
DB |||||
QY 1749 TGAATGCTGCTGATGAATTTTGAGAACACAGATGTTGATATGTTAGTTATAT 1808
DB |||||
QY 1983 TGGAGGGTATTGGCTATCTGCTGATGATATCAACAAATTTCTGATGGTTAGAT 2042
DB |||||
QY 1809 TGGAGGGTATTGGGTATTCTGCTGATGATATTAACAAATCTTTATGATGTAAGAT 1868
DB |||||
QY 2043 AGCTGAAGGAGAACATCCAAAGACATTCGGGAAAGTGACTATTTTACCCCAAGGAGA 2102
DB |||||
QY 1869 TGCTGAAGGTATCTGGCTGATGAATCAAAAGAAAGAGACTACTTTTACTGACCGAGGA 1928
DB |||||
QY 2103 ATTCGGTGTAGCAAGCAGATCCCTACTCTTGTGTAATTTGCTTATGATTAATGTC 2162
DB |||||
QY 1929 ATATAAGTGAATGAAGATGATCACTGGCAATGAAGAAATCTTTGATGTAAGTATC 1988
DB |||||
QY 2163 ATACTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCACCGAAC 2222
DB |||||
QY 1989 GTATCATAGATTCACGTGAA-----TTGTTTGGAGGTAGAGATGGTGTGATAGATTAG 2042
DB |||||
QY 2223 ACGTAATGCTGAGATTTGAATTAAGGACATTAATTCAAACATTTGGAAGAGCCCTTAC 2282
DB |||||
QY 2043 AAACCAAAATCCAGCAATGAAGTACCGAAATGAATGTTGTTGAAGAGCCCTTAC 2102
DB |||||
QY 2283 ATCAGAACACTGGCTTGTGTTAGGATATATAAGTAAGAACACCTGATAA 2330
DB |||||
QY 2103 ATCAGAAATTTGGATTGTGAGAAATTTTACAAAGTTTAAAGATTGATGAATA 2150
DB |||||

RESULT 15

US-10-417-375-92
; Sequence 92, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 3093

; TYPE: DNA
; ORGANISM: Mus musculus
US-10-417-375-92

Query Match 27.0%; Score 670.6; DB 18; Length 3093;
Best Local Similarity 58.9%; Pred. No. 3.9e-176;
Matches 1258; Conservative 0; Mismatches 799; Indels 78; Gaps 3;

QY 204 GCAGTGCCTCTCTCTCCCTTACCAATCCCTCTCTCTGCGCTGGCTTGGCGGCTTCAGCTCCGC 263
DB |||||
QY 153 GGACACACTTTAAAGACTTCTCATCTCTGCGATGCTGCTGTGTATCTTTTCTACTCG 212
DB |||||
QY 264 CTTCTTCCCGCTCATCCGCTTGAAGACATCATCCAGAGTTCGACCCGCTGTTTAACTA 323
DB |||||
QY 213 TCTTTTGTCTGCTGAGATTTGAAGTGTCTATCCATGATGATCCGTATTAATTA 272
DB |||||
QY 324 TAGATCAACACATCATCTTGCACTCTCATGGTCTCTATGAATTTTAAATGGTTTGAAGA 383
DB |||||
QY 273 TCGGACTACCGGTTTCTGGCTGAGAGGGGTTTATAAATTCATAAATGGTTTGAAGA 332
DB |||||
QY 384 AAGAGCATGGTATCCCACTAGGAAGATAGTAGTGCTACTGTTTACCCAGAGGTTGATCAT 443
DB |||||
QY 333 CCGGCTTGGTACCCCTTTGGGCGCAATCATTTGGAGGAACAAATTTACCCAGGTTTAAATCAT 392
DB |||||
QY 444 AACGCTGGCTTATTTTCATTTGATTTTAATATACATTTGAACATTAAGTTCACATAAGA 503
DB |||||
QY 393 CACTTCTGCTGCAATCTTACATGACTCCATTTCTTCCATATCATATTTGACATTCGGA 452
DB |||||
QY 504 CGTATGCTGCTTCTTCCCTGCAACCACTTTTAGCGGCTTACATCTATATCTACTTTCTCT 563
DB |||||
QY 453 TGCTGTGTTTCTGCGGCCCACTTTTCTCTCTTCCACCACTGTTAGTACCACCT 512
DB |||||
QY 564 TACAAGAGACTTTTGAACCAAGGAGCAGACTTTTACCTGCTGTTTATTTATGCTATGCT 623
DB |||||
QY 513 TACCAAGAGCTCAAGATGCGAGGCTGGCTTCTTCTGCTGCTGCAATGTTTCTGTAGT 572
DB |||||
QY 624 ACCAGGCTACATATCTCGGTGAGTAGCTGGATCCCTTTGATTAATGAAGCAATTTGCTAT 683
DB |||||
QY 573 TCTGGGTATATTTCTCGATCTGTAGCTGGCTCTATGATATGAAGAAATTTGCTATCT 632
DB |||||
QY 584 TGCACTTCAGTTCACATCTATTTATGGGTAATCTGTAAACCTGTAACCTGGCTGATTTT 743
DB |||||
QY 633 TTGATGCTGCTTACTTACTACATGCGATCAAGGAGTGAAGACTGGTTCATCTATG 692
DB |||||
QY 744 GACAAATGCTGCTGCTTATCTCTATTTCTATGCTCTCTGCTTGGGTTGTTATGAT 803
DB |||||
QY 693 GGCTGCCAAGTGTGCCCTCGCTTATTTCTACATGGTCTCTCATGGGAGGCTATGCT 752
DB |||||
QY 804 TATCATCAATCTTATTTCCATGCTGATTTTGTGTTGTACTGATGCGAGAGATACAGAA 863
DB |||||
QY 753 CCTGATCACTGATTTCTCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 812
DB |||||
QY 864 AAGAGTCTACATAGCATATAGCACTTTCTACATTTGTTGGGTTTAAATATTATCAATG 923
DB |||||
QY 813 CCGATCTACGTAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 872
DB |||||
QY 924 ACCTTTTGGGATTCAGCCCAATCAGAACAGTGAACACATGGCAGCTGCGAGGCTGCT 983
DB |||||
QY 873 TCTCTTGTGGTTCAGGCCGCTCTTTCTATCAAGAACATGCGAGGCTTTGGAGTGT 932
DB |||||
QY 984 TGCATTTGCTGCAAGCTTATGCTTTCTGCAATGCTGAGAGACCGAATTAACAAAACAAGA 1043
DB |||||
QY 933 TGGTCTCTGTGATCCATGCTTTCTGATGTTTACCTGCGCAGCAAGTTGAATCCACAGA 992
DB |||||
QY 1044 GTTCCAGACCTTTTCTTTTGGGTTATCATCACTAGCTGCTGCTGCTGCTGCTGCTGCT 1103
DB |||||
QY 993 ATTGGAAGTTCTTTTCCGAGTGTATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1052
DB |||||
QY 1104 CATCTATTGACTTATACAGGTTTACATTTGCAACCATGAGTGGCAGGTTTATTTCAATT 1163
DB |||||
QY 1053 AGCTCTCTCTCATGCTAACAGGAAAAATTTCTCCCTGGAAGGCGGCTTCTACTCTCTCT 1112
DB |||||
QY 1164 GGATCTGGGTATGCAAAAATACATTTCCATTTATGCAATGCTGCTGAGCATCAACC 1223
DB |||||

QY 420 TACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTCATTTGGATTTTAAATACATT 479
 Db 631 TACTGTTTACCAGGTTGATGATAACCGCTGGCTTATTCATTTGGATTTTAAATACATT 690
 QY 480 GAAATACACTGTTACATAGAGAGAGATGATGTTGTTCTTGGACCAACTTTAGGGGCT 539
 Db 691 GAACATACACTGTTACATAGAGAGAGATGATGTTGTTCTTGGACCAACTTTAGGGGCT 750
 QY 540 TACATCTATCTACTTCTCTGCTTACAGAGAACTTTGGAAACCAAGAGAGAGAGACTTTT 599
 Db 751 TACATCTATCTACTTCTCTGCTTACAGAGAACTTTGGAAACCAAGAGAGAGAGACTTTT 810
 QY 600 AGCTGCTTGTGTTTATGCTATGTTGACAGGCTACATATCTCGGTGAGTACGCTGCTT 659
 Db 811 AGCTGCTTGTGTTTATGCTATGTTGACAGGCTACATATCTCGGTGAGTACGCTGCTT 870
 QY 660 TGAATAGAGGCTATGCTATGTTGACAGGCTACATATCTCGGTGAGTACGCTGCTT 719
 Db 871 TGAATAGAGGCTATGCTATGTTGACAGGCTACATATCTCGGTGAGTACGCTGCTT 930
 QY 720 TGTAAAACTGGGTCAGTTTGTGACAAATGCTGCTGCTTATCTCTATTTCTATATGCT 779
 Db 931 TGTAAAACTGGGTCAGTTTGTGACAAATGCTGCTGCTTATCTCTATTTCTATATGCT 990
 QY 780 CTCTGCTTGGGCTGATGATTTATTCATCAATCTTATTCCTGCTGATTTGCTGCTT 839
 Db 991 CTCTGCTTGGGCTGATGATTTATTCATCAATCTTATTCCTGCTGATTTGCTGCTT 1050
 QY 840 GTTACTGATGACAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATGCT 899
 Db 1051 GGTACTGATGAC--AGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATGCT 1108
 QY 900 GGGTTT 905
 Db 1109 GGGTTT 1114
 RESULT 14
 US-10-032-585-6323
 ; Sequence 6323, Application US/10032585
 ; Publication No. US20030180953A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Terry, Roemer D.
 ; APPLICANT: Bo, Jiang
 ; APPLICANT: Charles, Boone
 ; APPLICANT: Howard, Bussey
 ; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 ; FILE REFERENCE: 10182-005-999-999/032,585
 ; CURRENT APPLICATION NUMBER: US/10/032,585
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 8000
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6323
 ; LENGTH: 2256
 ; TYPE: DNA
 ; ORGANISM: Candida albicans
 US-10-032-585-6323
 Query Match 27.8%; Score 689.2; DB 15; Length 2256;
 Best Local Similarity 59.9%; Pred. No. 1.9e-181;
 Matches 1250; Conservative 0; Mismatches 793; Indels 45; Gaps 4;
 QY 246 TGGCGGCTTACGCTCGGCTTCTCGCGCTATCCGCTTCCGAGCATCATCCAGGTT 305
 Db 105 TGCAGCTATTTCTCTGTTTATTTTCCGTTGATTCGATTTGAAGTATTTATTCATGAAT 164
 QY 306 CGACCGGTTTAACTATAGATCAACATCATCTTGTGATCTCATGCTTCTATGAAT 365
 Db 165 CGATCCTTGGTTCAATTTCCGAGCAACCAAAATATTTAGTCACTCATTCCTTTTATGAAT 224
 QY 366 TTTAAATTTGTTGATGAAGAGCATGCTATCCACTAGGAAGATAGTAGGTGGTACTGT 425

Db 225 TTTGAATTTGGTTTATGATAGAGAACTTTGGTACCATTGGGAAGAGTCACTGGTGTACTTT 284
 QY 426 TTACCAGGTTGATGATAACCGCTGGCTTATTCATTTGGATTTTAAATACATTGAA 482
 Db 285 ATATCCGGTTTAAATGGTGAATTTCAAGTCCCATTTGGCATATTTTACGTGATTTGCTTGC 344
 QY 483 CATACACTGTTACATAAGAGAGATGATGTTGTTCTTGCACCAACTTTTAGCGGCTTAC 542
 Db 345 CTTACCCTGTTGATTTAGAAATATTTGTTTATAGCAACGTTTCTCGGATTAAC 404
 QY 543 ATCTATATCTACTTCTCTGCTTACAGAGAACTTTGGAAACCAAGAGAGAGACTTTTACG 602
 Db 405 TGCATTTTGTACTTATTTTTCGACTAAAGAAATGAAGATTTCTAGTGCAGGATTTATTTGC 464
 QY 603 TGCTGTTTATGCTATTTGACAGGCTACATATCTCGGTGAGTACGCTGCTTCTTGA 662
 Db 465 AGCTATTTTATGCGGATTTGCCAGGTTATTTTCAAGATCAGTGGCTGGTCTTATGA 524
 QY 663 TAATGAAGGCTATGCTATTTTTCGACTTCACTGATTCATCACTATTTATGGAATAAATCTGT 722
 Db 525 TAATGAAGCAATTTGCCATTTACTTTTATTAATGCAACATTTTATTTCTGGATTAATCAAT 584
 QY 723 AAAAAGCTGGGTCAGTTTGTGACAAATGCTGCTGCTTATCTCTATTTCTATATGCTCTC 782
 Db 585 GAAATTTGGGTTTCACTTTCTATGCAATGCAATGACAGCATTTATTTCTATATGCTTAG 644
 QY 783 TGCTGGGTTGATGATTTATCACTATTTATCACTATTTATTTTCACTGATTTTGTGTTGT 842
 Db 645 TGCTGGGTTGATGATTTTTCATTTACCAATTTGATTTCCATTTACAGCTATTTTGTCTTGT 704
 QY 843 ACTGATGACAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGTTGG 902
 Db 705 TTTCAATGCTGTTTATTAATGCAAACTTTTACCTGTTTATCTACATGTTGCTTGGG 764
 QY 903 TTTAAATATCAATGACAGATACCTTTTGGGATTTCCAGCAATTCAGCAACAGAGTGAACA 962
 Db 765 TACTTTGGCATCAATGACAGATTTCCATTTGGGTTTTCACCAATTTAGATCAAAATGATCA 824
 QY 963 CATGGCAGCTGCGGTTGCTTTTGCATTTGCTGCAAGCTTATCTTTCTTGGCATTTCTGAG 1022
 Db 825 TATGGCTGATTTAGGATTTTGGATTTGTTACATTTAGTGGCTTTTGGTGTATGTTAA 884
 QY 1023 AGACCGATTAACAAACAGAGTTCCAGACCTTTTCTTTTGGGTTGATCACTAGTCTGC 1082
 Db 885 ATCAAAAGTTCCAAACCAAACTTTTAAATCACTTTGATAGTTTCCATTTGCTTGTGT 944
 QY 1083 AGTGCTGTTGCTTCTTATGTTGCTATCTATTTGCTTATACAGCTTATACAGTACCATGAG 1142
 Db 945 TGGATTTAGTATTTGGTGGATTTTGGATTTACAGCAATGGTGGATTTGCTTCTTGGAC 1004
 QY 1143 TGGCAGGTTTATTTCAATTTGGGATCTGCGGATGATGCAAAAAATACACATTTCAATTTTGC 1202
 Db 1005 AGGTAGATTTTATTTCTTATGGGATACAAATTTATGCCAAGATTTATTTCCAAATTTTGC 1064
 QY 1203 ATCAGTGTCTGAGCATCAACCTACGCTTGGGTTGCTTCTTCTTGTATCTACATTTCT 1262
 Db 1065 TTTGTTTCTGACATCAACCTACGCTTGGGTTGCTTCTTCTTGTATCTACATTTCT 1124
 QY 1263 TGTATGTAATCTTCCAGAGGCTTTGGTCTGATTTTGTGATCAAAATATCAACGATGAAGAGT 1322
 Db 1125 TATTTGGTATTTCCCGCTGGTATCTATTTATTTTCCAAAGATTTGAAGGATGAACAGT 1184
 QY 1323 ATTGTTGCTCTATATGCAATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1382
 Db 1185 TTTCAATATCAATTTACAGTGTATTTGTTTCTTATTTTGTGCTGCTGCTGCTGCTGCTGCTGCT 1244
 QY 1383 GTTGAATTTGACTTCCAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1442
 Db 1245 GTTGAATTTGACTTCCAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
 QY 1443 GCATCTATTTGGGAGGATGACATGAAAGGGAATAATCCACTGCTGGAGGACAGCAGTGA 1502
 Db 1305 TGTCTATTTGGACATTTG-----TTGATTTGTTCTACTGAGAAAGTTGGAAGATGA 1355

```
US-10-264-049-630
; Sequence 630, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,457
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 630
; LENGTH: 2660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2634)..(2634)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2637)..(2637)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-630

Query Match 36.0%; Score 992; DB 16; Length 2660;
Best Local Similarity 99.8%; Pred. No. 3.5e-238;
Matches 892; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1598 GGATTAGGCCCTTAATATAAAGCATTTGTCACCATGTTGATGCTGATGCTATTGATGAG 1647
DB 123 GGGTAGGCCCTTAATATAAAGCATTTGTCACCATGTTGATGCTGATGCTATTGATGAG 182
QY 1648 TTGCTGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1707
DB 183 TTGCTGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 242
QY 1708 GCCTCATACATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1767
DB 243 GCCTCATACATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
QY 1768 TGGCTAAGGCAAAATACAGATGAACATGACGAGTAAATGCTTGGTGGGATTAAGCTAT 1827
DB 303 TGGCTAAGGCAAAATACAGATGAACATGACGAGTAAATGCTTGGTGGGATTAAGCTAT 362
QY 1828 CAGATAGCTGGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGGAATACAG 1887
DB 363 CAGATAGCTGGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGGAATACAG 422
QY 1888 CACATAGCAGCTGGTGGGAAAGCTATGCTTCTTAATGAACACAGCAGCTTAAATATCAT 1947
DB 423 CACATAGCAGCTGGTGGGAAAGCTATGCTTCTTAATGAACACAGCAGCTTAAATATCAT 482
QY 1948 AGGACTCTAGATGATAGATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2007
DB 483 AGGACTCTAGATGATAGATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 542
QY 2008 GATGATATCAAAATTTCTCTGATGTTGATGATGATGATGATGATGATGATGATGATG 2067
DB 543 GATGATATCAAAATTTCTCTGATGTTGATGATGATGATGATGATGATGATGATGATG 602
QY 2068 ATTCGGGAAGTGAATTTTACCCACAGGAGAAATTCCTGTAGCAAAAGCAGGATCC 2127
DB 603 ATTCGGGAAGTGAATTTTACCCACAGGAGAAATTCCTGTAGCAAAAGCAGGATCC 662
QY 2128 CCTACTTTGTAATGCCCTTATGTAATAAATGTGTCATACAGATTTGGAGAAATGCAAG 2187
DB 663 CCTACTTTGTAATGCCCTTATGTAATAAATGTGTCATACAGATTTGGAGAAATGCAAG 722

US-10-296-115-629
; Sequence 629, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 629
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-629

Query Match 35.0%; Score 867.2; DB 16; Length 1114;
Best Local Similarity 98.8%; Pred. No. 1.7e-231;
Matches 895; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 1 ATGCGGAGCCCTCGGCCCGCGGAGCAAGCAAGTGTCTCAACTCGTCCCGCTGG 60
DB 211 ATGCGGAGCCCTCGGCCCGCGGAGCAAGCAAGTGTCTCAACTCGTCCCGCTGG 270
QY 61 AGTGGCTCTATGCTGCTGGAAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 271 AGTGGCTCTATGCTGCTGGAAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 330
QY 121 GCGCACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 179
DB 331 GCGCACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
QY 180 GGGGCTGTCGACGCGGCTGGTGGGAGTGGCTGCTCTCTTCCACATCTCTTCTTCCG 239
DB 391 GGGGCTGTCGACGCGGCTGGTGGGAGTGGCTGCTCTCTTCCACATCTCTTCTTCCG 450
QY 240 CTGGCTTGGCGGCTTCAGCTCGGCTCTTCCGCTCTTCCGCTCTTCCGCTCTTCCGCTCT 299
DB 451 CTGGCTTGGCGGCTTCAGCTCGGCTCTTCCGCTCTTCCGCTCTTCCGCTCTTCCGCTCT 510
QY 300 CGAGTTCGACCGGCTTAACTATAGATCAACATCATCTTGTGATCTTGTGCTTCTTCTT 359
DB 511 CGAGTTCGACCGGCTTAACTATAGATCAACATCATCTTGTGATCTTGTGCTTCTTCTT 570
QY 360 TGAATTTTAAATTTGGTTTGTGATGAAGAGCTGATCCACTAGGAAGAAATAGTAGGTGG 419
DB 571 TGAATTTTAAATTTGGTTTGTGATGAAGAGCTGATCCACTAGGAAGAAATAGTAGGTGG 630
```

QY 371 ATTGGTTTGTAAAGAGCATGGTATCCACTAGAGAAATAGTAGTGGTACTGTTTACC 430
 Db 282 ACTGGTTCGAGCGGCGATGGTATCCGCTCGGCGAGATGTGGCGGTACCGTCTATC 341
 QY 431 CAGGGTTCATGATACCGCTGGCTTATTCATGGATTTAAATACATTAAGACATACG 490
 Db 342 CCGGCTGATGATTAATACGTCGCGGGAATCCATTTGGCTGTCAGTACTCAACATACCG 401
 QY 491 TTACATAGAGAGATGTGTGTTCTTGCACCAACTTTTAGCGCCTTACATCTATAT 550
 Db 402 TCCATATTGCGATCTGCGTGTCTGCGCGGATCTTCAGTGGCTGACCTCATCT 461
 QY 551 CTACTTTCCTCTTACAGAGAACTTTGGAACCAAGAGCAGACCTTTAGCTGCTGTT 610
 Db 462 CCACCTACTCTGACCAAGAGCTGTGGTCCGGGCGCGGCTTTCGCGCGAGCT 521
 QY 611 TTATTGCTATTGTACCGGCTACATATCTCGGTGAGTGGATCTTTTGATTAATGAAG 670
 Db 522 TCATCGCCATCGTGGCTGATCATCAGTAGTGGTGGCTGGATCGTAGATTAACGAGG 581
 QY 671 GCATTGCTATTTCGCTTCAGTTCACATACATATTATGGGTAAATCTGTAAACATG 730
 Db 582 GCATTGCTATTTCGCTTCAGTTCACATACATATTATGGGTAAATCTGTAAACATG 641
 QY 731 GGTCAAGTTTTGGCAATGTGCTTATCTTATCTTATCTTATCTTATCTTATCTTAT 790
 Db 642 GATCGCTGTCTGGTGGCGCGAGCGCTTTGCTTACTTACTATCTTACTTACTTACT 701
 QY 791 GTGGTATTGATTTATCATCATCTTATTCACATGATGATTTGTTGTTACTGATGC 850
 Db 702 GTGGTATTGATTTATCATCATCTTATTCACATGATGATTTGTTGTTACTGATGC 761
 QY 851 AGAGATACGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGGGTAAAT 910
 Db 762 GCAGTACTCGCGGCTGTGTCAGCAGTACAGCACTTCTACATCTTGGGACTGCTGT 821
 QY 911 TATCAATCAGATACCTTTTGGGATTCAGCCCAATCAGAACATGAGTGAACATGGCAG 970
 Db 822 TCTCATGAGATCCCTTTCGTTGGGATTCAGCCCAATCAGAACATGAGTGAACATGGC 881
 QY 971 CTGAGGTGTCTTTGCATTTGCAAGCTTATGCTTTCTTGGAGTATCTGAGAGCCGAT 1030
 Db 882 CGCTGGGAGTGTGTTGCTCTTATGCGCGGCGACCTTTGCGCCATTTGCACTCGTGC 941
 QY 1031 TAAACAAACAGATTCAGACCTTTCTTTTGGGTATCTACATGCTGAGGTGCTG 1090
 Db 942 TGTCCGCAACAGTTCGGAAGCTGTTTCATGCTGGCGGATTTGCTGGGCGTTGGCG 1001
 QY 1091 TGTTCCTTAGTCTATTTAGTCTTATACAGTCTTACATTTGACCATGAGTGGCAGGT 1150
 Db 1002 TCTTTGCGCGTGTGCTGCTCACCATGCTGGCGGTGTGGCGCGTGGAGTGGAGCT 1061
 QY 1151 TTATTCATTTGGGATCTGGGTATGCAAAATACACATTTCCATTTATGATCAGTGT 1210
 Db 1062 TCTACTCGCTGGGATCTGGGTATGCAAAATACACATTTCCATTTATGATCAGTGT 1121
 QY 1211 CTGAGCATCAACCTACGATCTGGGTCTTTCTTTTGTGATCTACATTTCTTGTATGA 1270
 Db 1122 CGGAGCATCAGCCCAACCTTGGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1181
 QY 1271 CCTTCCAGAGCGCTTTGGTCTGCAATCAAAATATCAAGATGAAGATTTGTTG 1330
 Db 1182 CCTTCCAGAGCGCTTTGGTCTGCAATCAAAATATCAAGATGAAGATTTGTTG 1241
 QY 1331 CTCTATATGCAATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1390
 Db 1242 TGTCTGATCGCATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1301
 QY 1391 TGACCTCCAGTGTGCTGATCTGCTGCAATTTCTTCTTCTTCTTCTTCTTCTTCTT 1450
 Db 1302 TCACGCGGTTGTGCTGATCTGCGGAGTGGCGCTTTTCGGACTGTGGATGTTCTCC 1361
 QY 1451 TGGGGGATGA-----CATGAAAGGGGAAATCCACCTGTGGAG-----GACAGCAGTG 1498

Db 1362 TGCAAGAGATTCGTCTAAGCGAATGGGCACAGCCATTAAGCGCACCCAGCGATGATG 1421
 QY 1499 ATGAGGATGACAAAAGAAAACCAAGGAAATTTGTATGTAAGCGAGGTAAAGTGAAGAAAC 1558
 Db 1422 AAGCTGAGGATTCATTTGAGAGAGAGAGCCGTGTACGACAGGCTGGCAAGCTGAAGCATC 1481
 QY 1559 ATGCAACTGAAACAGGAAAAAACTGAAGAGGAGATTAAGGCCCTAATATAAAGACATTTGCA 1618
 Db 1482 GTACTAAGCATGATGCCAGCAGGATCTAGCGCTCAGCTCCCAACCTGAAGAGTATTGTTA 1541
 QY 1619 CCAATGTTGATGCTGATGCTATTGATGATGTTGCTGTCCACTGTACCTTGGGTCAACAACA 1678
 Db 1542 TTTTGGCGCTTCTAATGCTGTTGATGATGTTGCTGTCCACTGTGACGCTGGGTGACACGA 1601
 QY 1679 ATGCTACTCTAGTCCAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1735
 Db 1602 ATGCTACTCTAGTCCAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1661
 QY 1736 ATATCTAGATGATTTTAGAGAGAGCTTACTTTGGCTAAGGCAAAATACAGATGAACATG 1795
 Db 1662 ACATTTAGAGCATTTACAGAGAGCTTACTACTGCTTTTCGACAGACACTCCCGATGATG 1721
 QY 1796 CAGAGTAAATGCTTGGTGGGATTTAGGCTATCAGATAGCTGGATGGCTTAATAGAACTA 1855
 Db 1722 CTCGCGTATGCTTGGTGGGATTTAGGCTATCAGATAGCTGGATGGCTTAATAGAACTA 1781
 QY 1856 CGTTGGTGGATTAATACACTCGGAATAACAGCAATAGCACTGCTGGTGGGAAAAGCTATGT 1915
 Db 1782 CGTAGTGGATTAATACACTCGGAATAACAGCAATAGCACTGCTGGTGGGAAAAGCTATGT 1841
 QY 1916 CTTCTAATGAACAGCAGCGCTATAAATCATGAGAGCTCTAGATGTAGATTTATGTTTGG 1975
 Db 1842 CTTCAACGAGGAGAGAGTCTCTAGAAATATAGACATCTCTTGGAGTGGACTACGTTTGG 1901
 QY 1976 TTATTTTGGAGGGTATTGGCTATTTGGTGTATGATATCAACAAATTTCTCTGGATGG 2035
 Db 1902 TGATCTTGGCGTGTATCGCTATTTCTGGCGATGATATCAACAAATTTCTCTGGATGG 1961
 QY 2036 TTAGGATAGTGAAGAGAACTATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAC 2095
 Db 1962 TCCGAATTTGTGAGGAGAGAGTCCCAAGAGCATTAAGGAAAGGATTTTACCCCAAC 2021
 QY 2096 AGGAGAAATTCCTGTAGACAAAGCAGGATCCCTACTTTGTTGATTTGCTTATGTATA 2155
 Db 2022 GCGTGAATTCAGGATGATGCGGAGTGTCTCCGCGCTGCTCACTGCTTATGTACA 2081
 QY 2156 AATGTCTATACATACAGATTTGGAGAAATGCACTGATTTTGTACACCCCGAGTTTGG 2215
 Db 2082 AATTAAGCTACTACAGATTTGGGAAATTTGAAGTTTGGACTACAGAGGTCCATCTGGATATG 2141
 QY 2216 ACCGAAACAGTAAATGCTGAGATTTGGAATTAAGGACATTAATTTCAAAATTTGGAAGAAG 2275
 Db 2142 ATGCGACAGTAAACCGCTCAITTTGGGATTAAGGACTTCGATCTGACCTACCTGGAGAGG 2201
 QY 2276 CTTTACATCAGAACTGCTGCTTGTAGGATTAATTAAGTAAAGGACCTGATTAACAGGG 2335
 Db 2202 CCTACACCAAGAACTGCTGCTTGTTCGATCTATAGGTTGAAGAGCCGATGAGTTCA 2261
 QY 2336 AGACATTAAGTACAAACCTCGAGTCCCAACATTTTCCCAAAACAGAAAGTATTGTTCAA 2395
 Db 2262 ATAGCCATCAGTGAAGCCAGGAGAGAACGATTTCTCCAGCAACCTTCATTTTCGAGAA 2321
 QY 2396 AGAAGACTACAAAGAGAGAGGCTGCTACATTAATAAAGTGGTGTGTTTAAAGAAAGCA 2455
 Db 2322 AGAATCTTAAGCGTCCCAAG---GGCTACATACGAAACCGACCGGTTGTTGTTAAGGAA 2378
 QY 2456 AGAAATATCTAAGAGAGC 2474
 Db 2379 AACGACCTTGAATAAAC 2397

PRIOR APPLICATION NUMBER: PCT/US98/23435
PRIOR FILING DATE: 1998-11-04
PRIOR APPLICATION NUMBER: US 60/064,911
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,912
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,983
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,900
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: US 60/064,988
PRIOR FILING DATE: 1997-11-07
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 611
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 133
LENGTH: 1543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1055)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1143)
OTHER INFORMATION: n equals a,t,g, or c
US-10-621-401-133

Query Match 42.7%; Score 1060.4; DB 16; Length 1543;
Best Local Similarity 99.8%; Pred. No. 1.6e-285;
Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 267 CTTGCGCTGATCGCTTCGAAACATCCAGAGTTCGACCCGCTGTTAACTATAG 326
DB 1 CTTGCGCTGATCGCTTCGAAACATCCAGAGTTCGACCCGCTGTTAACTATAG 60
QY 327 ATCAACACATCATCTTCATCTCATGCTTCTATGAATTTTAAATTTGTTGATGAAG 386
DB 61 ATCAACACATCATCTTCATCTCATGCTTCTATGAATTTTAAATTTGTTGATGAAG 120
QY 387 AGCATGGTATCCATAGAGAAATAGTAGTGGTACTGTTTACCAGGTTGATGATAC 446
DB 121 AGCATGGTATCCATAGAGAAATAGTAGTGGTACTGTTTACCAGGTTGATGATAC 180
QY 447 CGCTGGCTTATTCATTGGATTTTAAATAATTAACATTAACATCTGTTTACATAAGACGT 506
DB 181 CGCTGGCTTATTCATTGGATTTTAAATAATTAACATTAACATCTGTTTACATAAGACGT 240
QY 507 ATGTGTGTTCTTCGACCACTTTTAGCGCTTACATCTATCTACTTCTGCTTAC 566
DB 241 ATGTGTGTTCTTCGACCACTTTTAGCGCTTACATCTATCTACTTCTGCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAGAGCAGACTTTTAGCTGCTGTTTATTGCTATGTACC 626
DB 301 AAGAGAACTTTGGAACCAAGAGCAGACTTTTAGCTGCTGTTTATTGCTATGTACC 360
QY 627 AGGCTACATATCTCGCTCAGTAGTGGATCCTTTGATTAATGAAGGATTCATTTTGC 686
DB 361 AGGCTACATATCTCGCTCAGTAGTGGATCCTTTGATTAATGAAGGATTCATTTTGC 420
QY 687 ACTTCAGTTTCACTACTATTTATGGGTAAATCTGTAAAACTGGGTCAAGTTTGGAC 746
DB 421 ACTTCAGTTTCACTACTATTTATGGGTAAATCTGTAAAACTGGGTCAAGTTTGGAC 480
QY 747 AATGTGCTGCTTATCTATTTCTATATGGTCTCTGCTGGGTGTTTATGATTTAT 806
DB 481 AATGTGCTGCTTATCTATTTCTATATGGTCTCTGCTGGGTGTTTATGATTTAT 540
QY 807 CATCAATCTTATTCACATGCAATGTTTGTGTTTACTGATCCAGATACAGCAAAAG 866
DB 541 CATCAATCTTATTCACATGCAATGTTTGTGTTTACTGATCCAGATACAGCAAAAG 600

QY 867 AGCTACATAGCATATAGCACCTTCTACATTTCTGGGTTTAAATATATCAATCCAGATACC 926
DB 601 AGCTACATAGCATATAGCACCTTCTACATTTCTGGGTTTAAATATATCAATCCAGATACC 660
QY 927 TTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTGCGAGTGTCTTTC 986
DB 661 TTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTGCGAGTGTCTTTC 720
QY 987 ATTGCTGCAAGCTTATGCTTTCTTGGGTATCTAGAGACCGATTAACAAAACAGAGCTT 1046
DB 721 ATTGCTGCAAGCTTATGCTTTCTTGGGTATCTAGAGACCGATTAACAAAACAGAGCTT 780
QY 1047 CCAGACCTTTCTTTTGGGTATCTAGAGACCGATTTCTTCTTCTTCTTCTTCTTCTTCT 1106
DB 781 CCAGACCTTTCTTTTGGGTATCTAGAGACCGATTTCTTCTTCTTCTTCTTCTTCTTCT 840
QY 1107 CTATTTGACTTATACAGGTTTACATTCACCATGGAGTGGAGGTTTATTCATTGTGGGA 1166
DB 841 CTATTTGACTTATACAGGTTTACATTCACCATGGAGTGGAGGTTTATTCATTGTGGGA 900
QY 1167 TACTGGGTATGCAAAAATACATTCCTCAATTTATGATCATGAGTCTGAGCATCAACCTAC 1226
DB 901 TACTGGGTATGCAAAAATACATTCCTCAATTTATGATCATGAGTCTGAGCATCAACCTAC 960
QY 1227 GACTTGGGTGCTTCTTCTTGGGTATCTAGAGACCGATTTCTTCTTCTTCTTCTTCTTCT 1286
DB 961 GACTTGGGTGCTTCTTCTTGGGTATCTAGAGACCGATTTCTTCTTCTTCTTCTTCTTCT 1020
QY 1287 TTGGTCTTGCATCAAAAATATCAAGATGAAGAGATTTGTT 1329
DB 1021 TTGGTCTTGCATCAAAAATATCAAGATGAAGAGATTTGTT 1063

RESULT 11
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1) ..(2417)
US-10-028-384-7

Query Match 38.5%; Score 955.8; DB 15; Length 2417;
Best Local Similarity 64.8%; Pred. No. 4.4e-256;
Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;

QY 191 AGCGCGCTGGGTGGAGTCCCTTCTCTCTTCCATCTCTTCTGCGCTGCGTGGCGG 250
DB 102 AGGTGGCTGGCTACAGAGCCTAAATCACCTTCGCTTCTGCTAAATCGCTGGCTGGCGG 161
QY 251 GCTTCAGCTCGCGCTCTTTCGCGCTCATCCGCTTCGAAGAGCATATCCAGAGTTCGACC 310
DB 162 GATTTCTCTCCCTCTTTCGCGCTCATCCGCTTCGAAGAGCATATCCAGAGTTCGATC 221
QY 311 CGTGGTTAACTATAGATCAACACATCATCTTCTGATCTCTCATGGTCTCTATGAATTTTAA 370
DB 222 CGTGGTTAACTATAGATCAACACATCATCTTCTGATCTCTCATGGTCTCTATGAATTTTAA 281

[illegible]

RESULT 9

```

RECORD 3
US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US2004018540A9
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a.t.g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-818-683-133

```

| | | |
|---|--|------|
| 241 | ATGTGTTTCCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTTCCTGCTTAC | 300 |
| 567 | AAGAGAACTTTGGAAACCAAGGACGAGACTTTTAGCTCTGTTTATTTAGTCTATTGTACC | 626 |
| 301 | AAGAGAACTTTGGAAACCAAGGACGAGACTTTTAGCTCTGTTTATTTGCTATTGTACC | 360 |
| 627 | AGGCTACATATCTCGGTGAGTCTGATCCCTTTGATAAATGAAGGCATTCGTAATTTTTCG | 686 |
| 361 | AGGCTACATATCTCGGTGAGTCTGATCCCTTTGATAAATGAAGGCATTCGTAATTTTTCG | 420 |
| 687 | ACTTCAGTTCCATACATATTTATGGGTAAATCTGTAAAAACTGGGTGAGTTTTTGGAC | 746 |
| 421 | ACTTCAGTTCCATACATATTTATGGGTAAATCTGTAAAAACTGGGTGAGTTTTTGGAC | 480 |
| 747 | AATGTCGTGCTGTTATCTTATTTCTATATGTCCTCTGTTGGGTGGTATGTAATTTAT | 806 |
| 481 | AATGTCGTGCTGTTATCTTATTTCTATATGTCCTCTGTTGGGTGGTATGTAATTTAT | 540 |
| 807 | CATCAATCTTATTCCTACTGCATGATTTGTGTGTTACTGATGAGAGATACAGCAAAAG | 866 |
| 541 | CATCAATCTTATTCCTACTGCATGATTTGTGTGTTACTGATGAGAGATACAGCAAAAG | 600 |
| 867 | AGTCTACATAGCATATAGCACCTTCTACATTTGGGTGTTAATATATTAATCAATGCAATACC | 926 |
| 601 | AGTCTACATAGCATATAGCACCTTCTACATTTGGGTGTTAATATATTAATCAATGCAATACC | 660 |
| 927 | TTTTGTGGGATTCGAGCCAATCAGAAACCAAGTGAAACACATGGCAGCTGCAGGTGCTTTTGC | 986 |
| 661 | TTTTGTGGGATTCGAGCCAATCAGAAACCAAGTGAAACACATGGCAGCTGCAGGTGCTTTTGC | 720 |
| 987 | ATTGCTGCAAGCTTATGCTTTCTTCTTGTCAGTATCTGAGAGACCGATTAACAAAAACAAGATT | 1046 |
| 721 | ATTGCTGCAAGCTTATGCTTTCTTCTTGTCAGTATCTGAGAGACCGATTAACAAAAACAAGATT | 780 |
| 1047 | CCAGACCTTTCTTTTGGGTGATACACTGCTGCAGGTGCTGTGTTTCCCTTAGTGTCAT | 1106 |
| 781 | CCAGACCTTTCTTTTGGGTGATACACTGCTGCAGGTGCTGTGTTTCCCTTAGTGTCAT | 840 |
| 1107 | CTATTTGCACATTATACAGGTATACATTGCACCATGGAGTGGCAGGTTTATTCATTGTGGGA | 1166 |
| 841 | CTATTTGCACATTATACAGGTATACATTGCACCATGGAGTGGCAGGTTTATTCATTGTGGGA | 900 |
| 1167 | TACTGGGTATGCAAAAATACATTCCTCAATTTATTCATCATGTCGTCTGAGCATCAACCTAC | 1226 |
| 901 | TACTGGGTATGCAAAAATACATTCCTCAATTTATTCATCATGTCGTCTGAGCATCAACCTAC | 960 |
| 1227 | GACTTGGGTGCTTTCTTTCTTTGATCTACATATCTTGTATGTATGTAACCTTCCAGCAGGCT | 1286 |
| 961 | GACTTGGGTGCTTTCTTTCTTTGATCTACATATCTTGTATGTATGTAACCTTCCAGCAGGCT | 1020 |
| 1297 | TGTTCTTCGATCAAAAATATCAAGATGAAGAGATTTCCTT | 1329 |
| 1021 | TGTTCTTCGATCAAAAATATCAAGATGAAGAGATTTCCTT | 1063 |
| RESULT 10 | | |
| US-10-621-401-133 | | |
| ; Sequence 133: Application US/10621401 | | |
| ; Publication No. US20040038277A1 | | |
| ; GENERAL INFORMATION: | | |
| ; APPLICANT: Rosen et al. | | |
| ; TITLE OF INVENTION: 125 Human Secreted Proteins | | |
| ; FILE REFERENCE: P2020P2C1 | | |
| ; CURRENT APPLICATION NUMBER: US/10/621,401 | | |
| ; CURRENT FILING DATE: 2003-07-18 | | |
| ; PRIOR APPLICATION NUMBER: US 09/974,879 | | |
| ; PRIOR FILING DATE: 2001-10-12 | | |
| ; PRIOR APPLICATION NUMBER: US 60/239,893 | | |
| ; PRIOR FILING DATE: 2000-10-13 | | |
| ; PRIOR APPLICATION NUMBER: US 09/818,683 | | |
| ; PRIOR FILING DATE: 2001-03-28 | | |
| ; PRIOR APPLICATION NUMBER: US 09/305,736 | | |
| ; PRIOR FILING DATE: 1999-05-05 | | |

PAGE 10

```

/ RESULT 10
/ US-10-621-401-133
/ : Sequence 133, Application US/10621401
/ : Publication No. US20040038277A1
/ : GENERAL INFORMATION:
/ : APPLICANT: Rosen et al.
/ : TITLE OF INVENTION: 125 Human Secreted Proteins
/ : FILE REFERENCE: F2020P2C1
/ : CURRENT APPLICATION NUMBER: US/10/621.401
/ : PRIOR FILING DATE: 2003-07-18
/ : PRIOR APPLICATION NUMBER: US 09/974,879
/ : PRIOR FILING DATE: 2001-10-12
/ : PRIOR APPLICATION NUMBER: US 60/239,893
/ : PRIOR FILING DATE: 2000-10-13
/ : PRIOR APPLICATION NUMBER: US 09/818,683
/ : PRIOR FILING DATE: 2001-03-28
/ : PRIOR APPLICATION NUMBER: US 09/305,736
/ : PRIOR FILING DATE: 1999-05-05

```


| | | | |
|----|------|--|------|
| QY | 387 | AGCATGGTATCCACTAGGAAGAAATAGTAGGTGGTACTGTTTACCAGGGTTGATGATAC | 446 |
| DB | 121 | AGCATGGTATCCACTAGGAAGAAATAGTAGGTGGTACTGTTTACCAGGGTTGATGATAC | 180 |
| QY | 447 | CGCTGGCCCTTATTCATTGGATTTTAAATACATTGAACATAACTGTTTCACATAAGACAGCT | 506 |
| DB | 181 | CGCTGGCCCTTATTCATTGGATTTTAAATACATTGAACATAACTGTTTCACATAAGACAGCT | 240 |
| QY | 507 | ATGTGTGTTTCCTTGACCAACCTTTAGCGGCCCTTACATCTATATCTACTTTCCTGCTTAC | 566 |
| DB | 241 | ATGTGTGTTTCCTTGACCAACCTTTAGCGGCCCTTACATCTATATCTACTTTCCTGCTTAC | 300 |
| QY | 567 | AAAGAGAACTTTGGAAACCAAGGACGAGCACTTTTAGCTGCTGTTTATTGCTATTGTACC | 626 |
| DB | 301 | AAAGAGAACTTTGGAAACCAAGGACGAGCACTTTTAGCTGCTGTTTATTGCTATTGTACC | 360 |
| QY | 627 | AGGCTACATATCTCGGTCAAGTAGCTGGATCCTTTGATATGAAGCAATGCTATTTTTGC | 686 |
| DB | 361 | AGGCTACATATCTCGGTCAAGTAGCTGGATCCTTTGATATGAAGCAATGCTATTTTTGC | 420 |
| QY | 687 | ACTTCAGTTTCACATACTATTTTATGGGTAAATCTGTAAAACTGGGTCAGTTTTTGAC | 746 |
| DB | 421 | ACTTCAGTTTCACATACTATTTTATGGGTAAATCTGTAAAACTGGGTCAGTTTTTGAC | 480 |
| QY | 747 | AATGTGCTGCTTATCCTATTTCTATATGTTCTCTGCTGGGTGGTATGATTTAT | 806 |
| DB | 481 | AATGTGCTGCTTATCCTATTTCTATATGTTCTCTGCTGGGTGGTATGATTTAT | 540 |
| QY | 807 | CATCAATCTTATTCACCTGCAGTATTTGTGTTTACTGATCGAGAGATACAGCAAAG | 866 |
| DB | 541 | CATCAATCTTATTCACCTGCAGTATTTGTGTTGTTACTGATCGAGAGATACAGCAAAG | 600 |
| QY | 867 | AGTCTACATAGCATATAGCACTTCTACATTTGSGTAAATATATCATCGACAGATACC | 926 |
| DB | 601 | AGTCTACATAGCATATAGCACTTCTACATTTGSGTAAATATATCATCGACAGATACC | 660 |
| QY | 927 | TTTTGTGGGAATTCAGCCAAATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTGC | 986 |
| DB | 661 | TTTTGTGGGAATTCAGCCAAATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTGC | 720 |
| QY | 987 | ATTGCTGCAAGCTTATGCTTTCTTGTCAGTATCTGAGAGACCGATTAAACAAACAGAGTT | 1046 |
| DB | 721 | ATTGCTGCAAGCTTATGCTTTCTTGTCAGTATCTGAGAGACCGATTAAACAAACAGAGTT | 780 |
| QY | 1047 | CCAGACCCCTTTCTTTTGGGTGTATCACTAGCTGCAGGTGCTGTTCCTTAGTGTAT | 1106 |
| DB | 781 | CCAGACCCCTTTCTTTTGGGTGTATCACTAGCTGCAGGTGCTGTTCCTTAGTGTAT | 840 |
| QY | 1107 | CTATTGTACTTTATACAGGTTACATTGCACCATGGAGTGGCAGGTTTTATTCATTCTGGGA | 1166 |
| DB | 841 | CTATTGTACTTTATACAGGTTACATTGCACCATGGAGTGGCAGGTTTTATTCATTCTGGGA | 900 |
| QY | 1167 | TACTGGGTATGCAAAATPACATTCCAATTTTGTCAATGAGTGTCTGAGCATCAACCTAC | 1226 |
| DB | 901 | TACTGGGTATGCAAAATPACATTCCAATTTTGTCAATGAGTGTCTGAGCATCAACCTAC | 960 |
| QY | 1227 | GACTTGGGTGCTTCTTCTTTTGATCTACATATCTTGTATGCTACCTCCACAGAGCCCT | 1286 |
| DB | 961 | GACTTGGGTGCTTCTTCTTTTGATCTACATATCTTGTATGCTACCTCCACAGAGCCCT | 1020 |
| QY | 1287 | TTGGTTTGCATCAAAATATCAACGATGAAGAGATTTGTT | 1329 |
| DB | 1021 | TTGGTTTGCATCAAAATATCAACGATGAAGAGATTTGTT | 1063 |

RESULT 8

| | | |
|-----|--|------|
| 601 | AGTCTACATAGCATATAGCACCTTTCTACATTGTGGGTTTAAATTATTATCAATCAGATACC | 660 |
| 601 | AGTCTACATAGCATATAGCACCTTTCTACATTGTGGGTTTAAATTATTATCAATCAGATACC | 660 |
| 927 | TTTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGAGCTGCAGAGTGCTCTTTCG | 986 |
| 927 | TTTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGAGCTGCAGAGTGCTCTTTCG | 986 |
| 661 | TTTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGAGCTGCAGAGTGCTCTTTCG | 720 |
| 661 | TTTTGTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGAGCTGCAGAGTGCTCTTTCG | 720 |
| 987 | ATTGTGCAAGCTTTATGCTTTCTTGAGATCTCTGAGAGACCGATTAAACAAAACAGAGTT | 1046 |
| 987 | ATTGTGCAAGCTTTATGCTTTCTTGAGATCTCTGAGAGACCGATTAAACAAAACAGAGTT | 1046 |

US-09-974-879-133

Query Match 42.7%; Score 1060.4; DB 10; Length 1543;
Best Local Similarity 99.8%; Pred. No. 1.6e-285;
Matches 1061; Conservative 0; Mismatches 2; Indels 0;

| | | | |
|------|----|--|------|
| 267 | QY | CTTGGCGGTCAATCCGCTTCGAAAGCAATCAATCCACAGAGTTTCGACCGGTGTTTAACTATAG | 326 |
| 1 | Db | CTTGGCGGTCAATCCGCTTCGAAAGCAATCAATCCACAGAGTTTCGACCGGTGTTTAACTATAG | 60 |
| 327 | QY | ATCAACACATCATCTTGCATCTCATGGGTTCATGAAATTTTAAATTCGTTTGAATGAAG | 386 |
| 51 | Db | ATCAACACATCATCTTGCATCTCATGGGTTCATGAAATTTTAAATTCGTTTGAATGAAG | 120 |
| 387 | QY | AGCATGGTATCCATAGGAAGAAATAGTAGTGGTACTGTTTACCCAGGGTTGATGATAAC | 446 |
| 121 | Db | AGCATGGTATCCATAGGAAGAAATAGTAGTGGTACTGTTTACCCAGGGTTGATGATAAC | 180 |
| 447 | QY | CGCTGGCCCTTAATCATTTGGATTTTAAATACATTTGAACATAACTGTTTCACATAAGACAGCT | 506 |
| 181 | Db | CGCTGGCCCTTAATCATTTGGATTTTAAATACATTTGAACATAACTGTTTCACATAAGACAGCT | 240 |
| 507 | QY | ATGTGTGTTCTTTCGACCAACTTTTACGGCCCTTACATCATATCTATCTTTCCTGCTTAC | 566 |
| 241 | Db | ATGTGTGTTCTTTCGACCAACTTTTACGGCCCTTACATCATATCTATCTTTCCTGCTTAC | 300 |
| 567 | QY | AAGAGAACTTTGGAAACCAAGAGCAGACATTTTAGCTGCTGTTTATTGCTATTGTACC | 626 |
| 301 | Db | AAGAGAACTTTGGAAACCAAGAGCAGACATTTTAGCTGCTGTTTATTGCTATTGTACC | 360 |
| 627 | QY | AGGTTACATATCTCGGTCACTAGCTGGATCCCTTTGATAATGAAGCAATGCTATTTTTCG | 686 |
| 361 | Db | AGGCTTACATATCTCGGTCACTAGCTGGATCCCTTTGATAATGAAGCAATGCTATTTTTCG | 420 |
| 687 | QY | ACTTCAGTTTCACATACTATTATTATGGGTGAAAACTGTFAAAACTGCGTCAGTTTTTTGGAC | 746 |
| 421 | Db | ACTTCAGTTTCACATACTATTATTATGGGTGAAAACTGTFAAAACTGCGTCAGTTTTTTGGAC | 480 |
| 747 | QY | AATGTGCTGCTTTATCCCTATTTCTATATGCTCTCTGCTGGGGTGGTTATGATTTTAT | 806 |
| 481 | Db | AATGTGCTGCTTTATCCCTATTTCTATATGCTCTCTGCTGGGGTGGTTATGATTTTAT | 540 |
| 807 | QY | CATCAATCTTATCCACTGCATGATTTTGTGTGTTTACTGTATGCAGAGATACAGACAAAG | 866 |
| 541 | Db | CATCAATCTTATCCACTGCATGATTTTGTGTGTTTACTGTATGCAGAGATACAGACAAAG | 600 |
| 867 | QY | AGTCTACATAGCATATAGCATCTTCTACATTTGCGGTTTAAATTTATCAATTCGAGATACC | 926 |
| 601 | Db | AGTCTACATAGCATATAGCATCTTCTACATTTGCGGTTTAAATTTATCAATTCGAGATACC | 660 |
| 927 | QY | TTTTGTGGATTTCCAGCCAAATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTTGC | 986 |
| 661 | Db | TTTTGTGGATTTCCAGCCAAATCAGAACAGTGAACACATGGCAGCTGCAGGTGCTTTTGC | 720 |
| 987 | QY | ATGTGCTGAAGCTTATGCTTTCTTGTGAGTATCTGAGAGCCGATTAACAAAACAAAGTT | 1046 |
| 721 | Db | ATGTGCTGAAGCTTATGCTTTCTTGTGAGTATCTGAGAGCCGATTAACAAAACAAAGTT | 780 |
| 1047 | QY | CCAGACCCCTTTTCTTTTTGGGTGATCCTAGCTGCAGGTGCTGTGTTTCTTTAGTGTCAT | 1106 |
| 781 | Db | CCAGACCCCTTTTCTTTTTGGGTGATCCTAGCTGCAGGTGCTGTGTTTCTTTAGTGTCAT | 840 |
| 1107 | QY | CTATTGTACTTATACAGGTTACATTTGCAATGGAGTGGCAGGTTTTTATTCATTTGGGA | 1166 |
| 841 | Db | CTATTGTACTTATACAGGTTACATTTGCAATGGAGTGGCAGGTTTTTATTCATTTGGGA | 900 |
| 1167 | QY | TACTGGGTATGCAAAATACACATTTCCAAATTTATGTCATCAGGTGCTGAGCATCAACCTAC | 1226 |
| 901 | Db | TACTGGGTATGCAAAATACACATTTCCAAATTTATGTCATCAGGTGCTGAGCATCAACCTAC | 960 |
| 1227 | QY | GACTTGGGTGCTTTTCTTTTGTGATCTACATATTCTTTGTATGTAAGCTTTCCAGCAGGCT | 1286 |
| 961 | Db | GACTTGGGTGCTTTTCTTTTGTGATCTACATATTCTTTGTATGTAAGCTTTCCAGCAGGCT | 1020 |

Qy 1287. TTGGTTCTGCATCAAAAAATATCAACGATGAAGAGTATTTGTT 1329
|||
Db 1021 TTGGTTCTGCATCAAAAAATATCAACGATGAAGAGTATTTGTT 1063
|||

RESULT 7

```

US-09-305-736-133
/ Sequence 133, Application US/09305736
/ Publication No. US20030088078A1
/ GENERAL INFORMATION:
/ APPLICANT: Feng et al.
/ TITLE OF INVENTION: 125 Human Secreted Proteins
/ FILE REFERENCE: P2020P1
/ CURRENT APPLICATION NUMBER: US/09/305,736
/ CURRENT FILING DATE: 1999-05-05
/ EARLIER APPLICATION NUMBER: PCT/US98/23435
/ EARLIER FILING DATE: 1998-11-04
/ EARLIER APPLICATION NUMBER: 60/064,911
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,912
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,983
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,900
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,988
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,987
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,908
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,984
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/064,985
/ EARLIER FILING DATE: 1997-11-07
/ EARLIER APPLICATION NUMBER: 60/066,094
/ EARLIER FILING DATE: 1997-11-17
/ EARLIER APPLICATION NUMBER: 60/066,100
/ EARLIER FILING DATE: 1997-11-17
/ EARLIER APPLICATION NUMBER: 60/066,089
/ EARLIER FILING DATE: 1997-11-17
/ EARLIER APPLICATION NUMBER: 60,066,095
/ EARLIER FILING DATE: 1997-11-17
/ EARLIER APPLICATION NUMBER: 60/066,090
/ EARLIER FILING DATE: 1997-11-17
/ NUMBER OF SEQ ID NOS: 612
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 133
/ LENGTH: 1543
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (1055)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (1143)
/ OTHER INFORMATION: n equals a,t,g, or c
/ US-09-305-736-133

```

```
Query Match      42.7%;   Score 1060.4;   DB 10;   Length 1543;
Best Local Similarity 99.8%;   Pred. No. 1.6e-285;
Matches 1061;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;
```

| | | | |
|-----|----|---|-----|
| 267 | QY | CTTCCGCGTCATCCGCTTGGAAAGCATCATCCACGAGTTCGACCCCGTGTTAACTATAG | 326 |
| 1 | DB | CTTCCGCGTCATCCGCTTGGAAAGCATCATCCACGAGTTCGACCCCGTGTTAACTATAG | 60 |
| 327 | QY | ATCAACACATCATCTTGCATCTCATGGGTTCTATGAATTTTAAATGGTTTGATCAAG | 386 |
| 61 | DB | ATCAACACATCATCTTGCATCTCATGGTTCATGAATTTTAAATGGTCATCAAG | 120 |

| | | | |
|----|------|--|------|
| QY | 1167 | FACTGGGTATGCAAAATACAGATTCCAAATTTATCGATCAGTGTCTGAGCATCAACCTAC | 1222 |
| Db | 901 | FACTGGGTATGCAAAATACAGATTCCAAATTTATCGATCAGTGTCTGAGCATCAACCTAC | 960 |
| QY | 1227 | GACTTGGGTGTCTTTCTTTCTTTGATCTACATATTCTTTGTATGTACCTTCCGAGCAGGCCT | 1286 |
| Db | 961 | GACTTGGGTGTCTTTCTTTCTTTGATCTACATATTCTTTGTATGTACCTTCCGAGCAGGCCT | 1020 |
| QY | 1287 | TGTTGTTCTGCAATCAAAATATCAACGATCAAGAGTATTGTT | 1329 |
| Db | 1021 | TGGTTCCTGCAATCAAAATATCAACGATCAAGAGTATTGTT | 1063 |

RESULT 6
 US-09-974-879-133
 Sequence 133, Application US/09974879
 Publication No. US20030028003A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 125 Human Secreted Proteins
 FILE REFERENCE: PZO2022
 CURRENT APPLICATION NUMBER: US/09/974,879
 CURRENT FILING DATE: 2001-10-12
 PRIOR APPLICATION NUMBER: US 60/239,893
 PRIOR FILING DATE: 2000-10-13
 PRIOR APPLICATION NUMBER: US 09/818,683
 PRIOR FILING DATE: 2001-03-28
 PRIOR APPLICATION NUMBER: US 05/305,736
 PRIOR FILING DATE: 1999-05-05
 PRIOR APPLICATION NUMBER: PCT/US98/23435
 PRIOR FILING DATE: 1998-11-04
 PRIOR APPLICATION NUMBER: US 60/064,911
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,912
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,983
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,900
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,988
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,987
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,908
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,984
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/064,985
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: US 60/066,094
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,100
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,089
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,095
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: US 60/066,090
 PRIOR FILING DATE: 1997-11-17
 NUMBER OF SEQ ID NOS: 611
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 133
 LENGTH: 1543
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (1055)
 OTHER INFORMATION: n equals a,t,g, or c
 NAME/KEY: SITE
 LOCATION: (1143)
 OTHER INFORMATION: n equals a,t,g, or c

QY 2128 CCTACTTTGTTGTAATGCTTATGATATAAATGTCATCTACAGATTGGGAGAAATGCAG 2187
 DB |||||
 QY 784 CCTACTTTGTTGTAATGCTTATGATATAAATGTCATCTACAGATTGGGAGAAATGCAG 843
 DB |||||
 QY 2188 CTGGATTTTCGTACACCCCGAGCTTTTGACCGAACAACGTAATGCTGAGATTGGAAATAG 2247
 DB |||||
 QY 844 CTGGATTTTCGTACACCCCGAGCTTTTGACCGAACAACGTAATGCTGAGATTGGAAATAG 903
 DB |||||
 QY 2248 GACATTAAATTCACAAATTTGGAAGAGCTTTTACATCAGACACGCTGCTTTAGGATA 2307
 DB |||||
 QY 904 GACATTAAATTCACAAATTTGGAAGAGCTTTTACATCAGACACGCTGCTTTAGGATA 963
 DB |||||
 QY 2308 TATAAAGTAAAGCACTGATACAGGAGAGACATTAGATCAAAACCTCGAGTCACCAAC 2367
 DB |||||
 QY 964 TATAAAGTAAAGCACTGATACAGGAGAGACATTAGATCAAAACCTCGAGTCACCAAC 1023
 DB |||||
 QY 2368 ATTTTCCCAAAACAGAGTATTTGTCARAAGAGACTTACCMAAGAGAGCGGTGCTACATT 2427
 DB |||||
 QY 1024 ATTTTCCCAAAACAGAGTATTTGTCARAAGAGACTTACCMAAGAGAGCGGTGCTACATT 1083
 DB |||||
 QY 2428 AAAAATAAGCTGCTTTTAAAGAAAGCAAGAAATATCTAAGAGACTGTTTAA 2481
 DB |||||
 QY 1084 AAAAATAAGCTGCTTTTAAAGAAAGCAAGAAATATCTAAGAGACTGTTTAA 1137
 DB |||||
 RESULT 4
 US-10-106-698-330
 ; Sequence 330, Application US/10106698
 ; Publication No. US20030103690A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: P0005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: Patent In Ver. 3.0
 ; SEQ ID NO 330
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1203)..(1203)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-106-698-330
 Query Match 42.8%; Score 1061; DB 15; Length 1209;
 Best Local Similarity 99.8%; Pred. No. 9.5e-286;
 Matches 1061; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 267 CTTCCGCGTCATCCGCTTCGAAAGCATCATCCACGAGTTCCGCGTGTAACTATAG 326
 DB |||||
 QY 327 ATCAACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATTTGGTTGATGAAG 386
 DB |||||
 QY 61 ATCAACACATCATCTTGGCATCTCATGGTTCATGAATTTTAAATTTGGTTGATGAAG 120
 DB |||||
 QY 387 AGCATGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 446
 DB |||||
 QY 121 AGCATGTATCCACTAGGAAGATAGTAGTGGTACTGTTTACCAGGGTTGATGATAAC 180
 DB |||||
 QY 447 CGCTGGGCTTATTCATTGGATTTTAAATACATTGAACTAACTGTTTACATAAGAGAGCT 506
 DB |||||
 QY 181 CGCTGGGCTTATTCATTGGATTTTAAATACATTGAACTAACTGTTTACATAAGAGAGCT 240
 DB |||||

QY 507 ATGTGTGTTCTCTTGACCAACCTTTTAGCGGCTTACATCTATATCTACTTTCTGCTTAC 566
 DB |||||
 QY 241 ATGTGTGTTCTCTTGACCAACCTTTTAGCGGCTTACATCTATATCTACTTTCTGCTTAC 300
 DB |||||
 QY 567 AAGAGAACTTTGGAAACCAAGAGAGAGACTTTTACGCTGCTTTTATGCTATGTTAC 626
 DB |||||
 QY 301 AAGAGAACTTTGGAAACCAAGAGAGAGACTTTTACGCTGCTTTTATGCTATGTTAC 360
 DB |||||
 QY 627 AGGCTACATATCTCCGTCAGTAGCTGATCCTTTGATAAAGAGGCAATTCGTTATTTTC 686
 DB |||||
 QY 361 AGGCTACATATCTCCGTCAGTAGCTGATCCTTTGATAAAGAGGCAATTCGTTATTTTC 420
 DB |||||
 QY 687 ACTTCAGTTTCACATCTACTTTTATGGGTAATACTGTAAAACTGGGTCAGTTTTCGAC 746
 DB |||||
 QY 421 ACTTCAGTTTCACATCTACTTTTATGGGTAATACTGTAAAACTGGGTCAGTTTTCGAC 480
 DB |||||
 QY 747 ATGTGCTGCTGCTTATCCTATTTCTATATGCTCTGCTTGGGTCGTTATGTTATTTAT 806
 DB |||||
 QY 481 ATGTGCTGCTGCTTATCCTATTTCTATATGCTCTGCTTGGGTCGTTATGTTATTTAT 540
 DB |||||
 QY 807 CATCAATCTTATTCACCTGCTATTTTGTGTTGTTACTGTATGATCAGAGATACAGCAAAAG 866
 DB |||||
 QY 541 CATCAATCTTATTCACCTGCTATTTTGTGTTGTTACTGTATGATCAGAGATACAGCAAAAG 600
 DB |||||
 QY 867 AGTCTACATAGCATATAGCACTTTCTACATTTGCGGTTTATATATTCATTCAGATAC 926
 DB |||||
 QY 601 AGTCTACATAGCATATAGCACTTTCTACATTTGCGGTTTATATATTCATTCAGATAC 660
 DB |||||
 QY 927 TTTTGTGGGATTCAGCCCAATCAGAAAGTGAACACATGCGAGCTGCGAGGTGCTTTTC 986
 DB |||||
 QY 661 TTTTGTGGGATTCAGCCCAATCAGAAAGTGAACACATGCGAGCTGCGAGGTGCTTTTC 720
 DB |||||
 QY 987 ATGTCTGCAAGCTTATGCTTTCTGTCAGTATCTGAGAGCCGATTAACAAACAGAGTT 1046
 DB |||||
 QY 721 ATGTCTGCAAGCTTATGCTTTCTGTCAGTATCTGAGAGCCGATTAACAAACAGAGTT 780
 DB |||||
 QY 1047 CCAGAGCCCTTTTCTTTTGGGTGATCAGTACAGCTGCGAGGTGCTGTTCTTATGTTAT 1106
 DB |||||
 QY 781 CCAGAGCCCTTTTCTTTTGGGTGATCAGTACAGCTGCGAGGTGCTGTTCTTATGTTAT 840
 DB |||||
 QY 1107 CTATTTGACTTATACAGTTTACATTTGACAGTGGAGTGGAGGTTTATTCATTTGGGA 1166
 DB |||||
 QY 841 CTATTTGACTTATACAGTTTACATTTGACAGTGGAGTGGAGGTTTATTCATTTGGGA 900
 DB |||||
 QY 1167 TACTGGGTATGCAAAATATACATTTCCAAATTTATTCATGCTGCTGAGCATCAACCTAC 1226
 DB |||||
 QY 901 TACTGGGTATGCAAAATATACATTTCCAAATTTATTCATGCTGCTGAGCATCAACCTAC 960
 DB |||||
 QY 1227 GACTTGGGTGCTTTCTTTTGTGATCTACATTTCTTGTATGATACCTTCCAGAGGCT 1286
 DB |||||
 QY 961 GACTTGGGTGCTTTCTTTTGTGATCTACATTTCTTGTATGATACCTTCCAGAGGCT 1020
 DB |||||
 QY 1287 TTGGTCTGCTCAAAATATCAACGATGAAAGAGTATTTGTT 1329
 DB |||||
 QY 1021 TTGGTCTGCTCAAAATATCAACGATGAAAGAGTATTTGTT 1063
 DB |||||
 RESULT 5
 US-10-264-237-412
 ; Sequence 412, Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P013P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: Patent In Ver. 3.1


```

Db 2281 ACATCAGAACCTGGCTGTGTAGGATATATAAAGTAAAGACCTGATAACAGGAGACA 2340
Qy 2341 TTAGATCACAACCTCGAGTCACCAATTTCCCAAAACAGAGTATTTGTCAAGAG 2400
Db 2341 TTAGATCACAACCTCGAGTCACCAATTTCCCAAAACAGAGTATTTGTCAAGAG 2400
Qy 2401 ACTACCAAAAGGAGCGTGGCTGCTACATTAATAATAGCTGGTTTAAAGAAAGGCAAGAAA 2460
Db 2401 ACTACCAAAAGGAGCGTGGCTGCTACATTAATAATAGCTGGTTTAAAGAAAGGCAAGAAA 2460
Qy 2461 ATATCTAAGAGAGCTGTTTAA 2481
Db 2461 ATATCTAAGAGAGCTGTTTAA 2481

RESULT 2
US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

Query Match 83.8%; Score 2078.6; DB 15; Length 2710;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 234; Indels 9; Gaps 1;

Qy 1 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTGTCCCTCAACTCGTCCCGCTGG 60
Db 77 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTGTCCCTCAACTCGTCCCGCTGG 136
Qy 61 AGTGGCTCATGCGCTCGGGAACAGCGCGGCAAGCGGCGGCGGCGGCGGCGGCGG 120
Db 137 AGCGGCTCATGCGCTCGGGAACAGCGCGGCAAGCGGCGGCGGCGGCGGCGGCGG 196
Qy 121 GGCACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 197 GCGTCCAGGCGG-----GGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 247
Qy 181 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
Db 248 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 307
Qy 241 TGGCTGCGGCGGCTCAGCTCGGCGCTCTCGCGGTCATCGGCTCGAAGCATCATCCAC 300
Db 308 TGGCTGCGGCGGCTCAGCTCGGCGCTCTCGCGGTCATCGGCTCGAAGCATCATCCAC 367
Qy 301 GAGTTCGACCCGCTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCTAT 360
Db 368 GAGTTCGACCCGCTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCTAT 427
Qy 361 GAATTTTAAATGCTTGTGTAAGAGAGCATGTAATCCACTAGGAGAGTATAGTAGTGGT 420
Db 428 GAGTTTCTAATTTGTTTGTATGAAGAGCATGTAATCCACTAGGAGAGTATAGTAGTGGT 487
Qy 421 ACTGTTTACCCAGGTTGATGATAACCGCTGSCCTTATTCATGTTGATTTTAAATACATTG 480

```

```

Db 488 ACCGTTTACCAGGCTGTATGATAACAGCTGGCTTATTCATTGGATTTTAAATACATTG 547
Qy 481 AACATAAGCTGTTTACATAAGAGAGACGTATGTGTGTTTCCCTTGACCAACTTTTTCAGGGCTT 540
Db 548 AACATAAGCTGTTTACATAAGAGAGAGTGTGTGTATTTCCCTTGACCAACTTTTTCAGGGCTT 607
Qy 541 ACATCTATATCTTCTCTCTTACAAAGAGAACTTTTGGAAACCAAGAGAGAGAGCTTTTAA 600
Db 608 ACATCCATATCTTCTCTCTTACAAAGAGAACTTTTGGAAACCAAGAGAGAGAGCTTTTAA 667
Qy 601 GCTGCTGTTTATTTGCTTATTTACAGGCTACATATCTCGGTCAGTAGCTGATCTCTTT 660
Db 668 GCTGCTGTTTATTTGCTTATTTACAGGCTACATATCTCGGTCAGTAGCTGATCTCTTT 727
Qy 661 GATAATGAAGGCAATTTGCTTATTTTGCATCTCAGTTTCACTACTTATTTATGGTAAATCT 720
Db 728 GATAATGAAGGCAATTTGCTTATTTTGCATCTCAGTTTCACTACTTATTTATGGTAAATCT 787
Qy 721 GTAAAAAATGCGTCTGAGTTTGTGGACAATGTGCTGCTTATCTCTTCTATATATGCTC 780
Db 788 GTGAAGACCGGCTGTGTCTGTGACAAATGTGCTGCTGTGTCTCATATTTCTACATGCTC 847
Qy 781 TCTGCTTGGGTGGTATGATTTATCATCAATTTATTCATCTTACCTGCAATGATTTGTTG 840
Db 848 TCTGCTTGGGTGGTATGATTTATCATCAATTTATTCATCTTACCTGCAATGATTTGTTG 907
Qy 841 TTACTGATGACAGATACAGCAAAAGAGTCTACATAGCATATGACCTTTCTCATATGCTG 900
Db 908 CTGCTGATGACAGAGGTACAGCAAGAGTCTACATAGCATATGACCTTTCTCATATGCTG 967
Qy 901 GGTTTAATATTTATCAATGCAGATACCTTTTGTGGATTTCCAGCAATTCAGAAACAGTGAA 960
Db 968 GGTTTAATATTTATCAATGCAGATACCTTTTGTGGATTTCCAGCAATTCAGAAACAGTGAA 1027
Qy 961 CACATGGCAGCTGACAGGTGCTTTTGCATTTGCTGCAAGCTTATGCTTCTTCTGAGTATCTG 1020
Db 1028 CACATGGCAGCTGACAGGTGCTTTTGCCTGCTGCAAGCTTATGCTTCTTCTGAGTATCTG 1087
Qy 1021 AGAGACCGATTTAAACAAAACAGAGTTCCAGACCTTTTCTTTTGGGTGATCACTAGCT 1080
Db 1088 AGAGACCGGTTTGAACAAAACAGAGTTCCAGACCTTTTCTTTTGGGTGATCACTAGCT 1147
Qy 1081 GAGGTGCTGCTGCTTCTTCTTGTGCTCATCTTATTTGATTTATACAGTTACATTTGACCATGG 1140
Db 1148 GAGGTGCTGCTGCTTCTTCTTGTGCTCATCTTATTTGATTTATACAGTTATTTGACCATGG 1207
Qy 1141 AGTGGCAGGTTTATTTCTATTTGCGGATCTGGGTATGCAAAAATACACATTTCAATTTAT 1200
Db 1208 AGTGGCAGGTTTATTTCTATTTGCGGATCTGGGTATGCAAAAATACACATTTCAATTTAT 1267
Qy 1201 GCATCAGTCTCTGACATCAACCTACGATTTGGGTGCTTCTTCTTGTGATCTACATATT 1260
Db 1268 GCATCAGTCTCTGACATCAACCTACGATTTGGGTGCTTCTTCTTGTGATCTACATATT 1327
Qy 1261 CTGTTATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1320
Db 1328 CTGTTATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1387
Qy 1321 GTATTTGTTGCTTCTATATGCAATCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
Db 1388 GTATTTGTTGCTTCTATATGCAATCAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1447
Qy 1381 ATGTTGACTTTGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1440
Db 1448 ATGTTGACTTTGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1507
Qy 1441 GAGCACTATTTGGGGGATGACATGAAAGGCAAAATCCACTGTTGGAGGACAGCATGAT 1500
Db 1508 GAGCACTATTTGGGGGATGACATGAAAGGCAAAATCCACTGTTGGAGGACAGCATGAT 1567
Qy 1501 GAGGATGACAAAAGAAACCAAGGAAATTTGTATAGGAGGAGTAAAGTGAAGAACAT 1560

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 07:01:53 ; Search time 1239.99 Seconds
(without alignments)
11031.546 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcgagacccctggcccc.....tatctaagaagactgtttaa 2481

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2755760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2481 | 100.0 | 2481 | 15 | US-10-028-384-1 |
| 2 | 2078.6 | 83.8 | 2710 | 15 | US-10-028-384-3 |
| 3 | 1114.8 | 44.9 | 1828 | 10 | US-09-945-527-62 |
| 4 | 1061 | 42.8 | 1209 | 15 | US-10-106-698-330 |
| 5 | 1061 | 42.8 | 1209 | 16 | US-10-264-237-412 |
| 6 | 1060.4 | 42.7 | 1543 | 10 | US-09-974-879-133 |
| 7 | 1060.4 | 42.7 | 1543 | 10 | US-09-305-736-133 |
| 8 | 1060.4 | 42.7 | 1543 | 10 | US-09-818-683-133 |
| 9 | 1060.4 | 42.7 | 1543 | 11 | US-09-818-683-133 |
| 10 | 1060.4 | 42.7 | 1543 | 16 | US-10-621-401-133 |
| 11 | 955.8 | 38.5 | 2417 | 15 | US-10-028-384-7 |
| 12 | 892 | 36.0 | 2660 | 16 | US-10-264-049-630 |

| | | | | | | |
|----|-------|------|------|----|----------------------|--------------------|
| 13 | 867.2 | 35.0 | 1114 | 16 | US-10-296-115-629 | Sequence 629, App |
| 14 | 689.2 | 27.8 | 2256 | 15 | US-10-032-585-6323 | Sequence 6323, App |
| 15 | 670.6 | 27.0 | 3093 | 18 | US-10-417-375-92 | Sequence 92, Appl |
| 16 | 670.6 | 27.0 | 3094 | 15 | US-10-028-384-9 | Sequence 9, Appl |
| 17 | 646.6 | 26.1 | 2472 | 15 | US-10-171-581-112 | Sequence 112, App |
| 18 | 646.6 | 26.1 | 2472 | 15 | US-10-028-384-11 | Sequence 11, Appl |
| 19 | 646.6 | 26.1 | 2472 | 15 | US-10-172-118-742 | Sequence 742, App |
| 20 | 646.6 | 26.1 | 2472 | 16 | US-10-342-887-742 | Sequence 742, App |
| 21 | 646.6 | 26.1 | 3046 | 18 | US-10-417-375-95 | Sequence 95, Appl |
| 22 | 638 | 25.7 | 5404 | 18 | US-10-417-375-99 | Sequence 99, Appl |
| 23 | 638 | 25.7 | 5404 | 18 | US-10-417-375-97 | Sequence 97, Appl |
| 24 | 616.2 | 24.8 | 2839 | 18 | US-10-425-115-150745 | Sequence 150745 |
| 25 | 610.2 | 24.6 | 2881 | 16 | US-10-425-114-14408 | Sequence 14408, A |
| 26 | 609.4 | 24.6 | 2866 | 16 | US-10-320-797-2305 | Sequence 2305, Ap |
| 27 | 606.2 | 24.4 | 2232 | 15 | US-10-128-714-7139 | Sequence 7139, Ap |
| 28 | 585.2 | 23.6 | 2779 | 17 | US-10-437-963-99904 | Sequence 99904, A |
| 29 | 539.4 | 21.7 | 2157 | 9 | US-09-801-368-387 | Sequence 387, App |
| 30 | 539.4 | 21.7 | 2157 | 18 | US-10-793-639-318 | Sequence 318, App |
| 31 | 537.8 | 21.7 | 2733 | 15 | US-10-028-384-5 | Sequence 5, Appl |
| 32 | 515 | 20.8 | 1848 | 15 | US-10-128-714-2139 | Sequence 2139, Ap |
| 33 | 500 | 20.2 | 500 | 9 | US-09-988-588-1643 | Sequence 1643, Ap |
| 34 | 458.2 | 18.5 | 558 | 14 | US-10-052-283-433 | Sequence 433, App |
| 35 | 456.4 | 18.4 | 1969 | 15 | US-10-128-714-1139 | Sequence 1139, Ap |
| 36 | 456.4 | 18.4 | 2603 | 15 | US-10-128-714-6139 | Sequence 6139, Ap |
| 37 | 456.4 | 18.4 | 3969 | 15 | US-10-128-714-139 | Sequence 139, App |
| 38 | 456.4 | 18.4 | 4603 | 15 | US-10-128-714-5139 | Sequence 5139, Ap |
| 39 | 450.4 | 18.2 | 3141 | 18 | US-10-425-115-130787 | Sequence 130787 |
| 40 | 443.4 | 17.9 | 485 | 10 | US-09-818-995-11283 | Sequence 11283, A |
| 41 | 430 | 17.3 | 430 | 9 | US-09-878-178-717 | Sequence 717, App |
| 42 | 430 | 17.3 | 430 | 13 | US-10-046-935-717 | Sequence 717, App |
| 43 | 430 | 17.3 | 430 | 14 | US-10-146-502-717 | Sequence 717, App |
| 44 | 429 | 17.3 | 3197 | 16 | US-10-424-599-111541 | Sequence 111541 |
| 45 | 415.4 | 16.7 | 616 | 9 | US-09-879-536-332 | Sequence 332, App |

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

| | | | | |
|-----------------------|--------------|---|------------|--------------|
| Query Match | 100.0% | Score 2481; | DB 15; | Length 2481; |
| Best Local Similarity | 100.0% | Pred. No. 0; | | |
| Matches 2481; | Conservative | 0; | Mismatches | 0; |
| | | | Indels | 0; |
| | | | Gaps | 0; |
| QY | 1 | ATGGCGGAGCCCTGGGCCCCGGAGAGCAAGCAAGTCCTCACTGCTCCCGTGG | 60 | |
| DB | 1 | ATGGCGGAGCCCTGGGCCCCGGAGAGCAAGCAAGTCCTCACTGCTCCCGTGG | 60 | |
| QY | 61 | AGTGGCCCTATGGCCCTGGGAAAACAGCCGGCAACGCGGCGCCGGGCCCCAGTGC | 120 | |
| DB | 61 | AGTGGCCCTATGGCCCTGGGAAAACAGCCGGCAACGCGGCGCCGGGCCCCAGTGC | 120 | |

| | | | |
|------|----|--|------|
| 1604 | QY | TAATAAGCATTTGTCACACAGTGTGATGCTGATGCTATTTGATGATGTTTGTCTGTCCACTGTA | 1663 |
| 140 | DB | TGAAGAGATTGTTTATTTTGGCCGGTCTTAATGCTGTTGATGATGTTGCTGTGCCACTGCA | 81 |
| 1664 | QY | CCTGGGTCCAAAGCAATGCGCTACTCTAGTGCCAAAGTGTAAGTCCTGGCCTCATACAA | 1718 |
| 80 | DB | CGTGGGTGACACAGCAATATGCGCTACTCCAGTCCCTCCCAATGTTCTTGGCTTTCCACAA | 26 |

RESULT 15

RESULTS IS
US-09-270-767-16600/c

US-03-270#767-16800/C
: sequence 16600. Application US/09270767: sequence 16600, APM
: Patent No. 6703491

; PATENT NO. 6703491
; GENERAL INFORMATION.

ADJUTANT: HONORARY CHIEF OF POLICE

APPLICANT: Homburger et al.

10; TITLE OF INVENTION: Nucleic acids and pro-

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/0

; CURRENT FILING DATE: 1999-03

; NUMBER OF SEQ ID NOS: 62517

: SOFTWARE: Paten

: SOF IWARE: 1660

; SEQ ID NO 166
; LENGTH: 535; LENGTH: 560
; TYPE: TXT

TYPE: DNA

Query Match 8:1%; Score 201.4; DB 4; Length 560;
Best Local Similarity 63.9%; Pred. No. 2.9e-44;
Matches 342; Conservative 0; Mismatches 181; Indels 12; Gaps 2;

Qy 1196 TTATTGCATCATGTCGTGAGCATCAACTACGACTTGGGTGCTTTCTTCTTTGATCTAC 1255

Db 560 TCATTGCATCCGTGTGGAGCATCAGCCCACTTGGTTCGTCTCTTCTGATCTGC 501

1256 ATATTCTGTATGTACCTTCCAGCAGGCCTTTGGTTCTGCATCAAAAAATCAACGATG 1315
Qy
500 ACATCATGGTGTGCGCCTTCCAGTGGGAGTGTGGTACTGCATCAGCAGATCAACGACG 441
Db

1316 AAAGAGTATTTCCTGCTCTATATGCAATCAGTGTCTCTACTTTCTGTGGAGTATGGTGC 1375

440 AGCGCGTTTTCTGTGGTCTGTACGCCAATCAGTGTGGGTTTACTTTCGTGGTGTATGGTGC 381

Qy 1376 GACTGATGTTGACTTTGACTCCAGTCGTGTGATGCTGCTGCAATGGCCCTTTTCAATG 1435

Db 380 GTTGTGTTGACCTTACGCGGGTGTGTGATGCTGGCGGAGTGGCCCTTTTCGGGAC 321

Qy 1436 TTTTGGACCATTTGGGGATG-----CATGAAAGGGAAATCCACTGTGGAG 1489

Db 320 TGTGGATGTGTCTCGAAGAGGATTCGTCTAAGCGAATGGGCACAGCCATAAGCGCAG 261

Qy 1489 -----GACAGCAGTGATGAGGATGACAAAGAACCCAGGAATTTCTGTATGATAAGSCAG 1543

Db 260 CCACCGAGTGATGAGCTGAAGTGGATTCCATTGAGAGAAGACGCTGTACGACAAAGGCTG 201

Qy 1544 GTAAAGTGAAGAAACATGCCTCACTGAACAGGAAAAAAGTGAAGAGGGATTAGGCCCTAATA 1603

Db 200 GCAAGCTGAAGCATCGTACTAAGCATGATGCCAGCAGGATACTGGCGTCAGCTCCACCC 141

Qy
1604 TAAAAAGCATTGCACCATGTTGATGCTGATGCTATTGATGATGTTGCTGCTCCACCTGA 1669
Dbb
140 TCAAGAGTATTGTTATTTGGCGGTTCTAATGCTGTTGATGATGTTGCTGCTCCACCTGCA 81

QY
 1664 CTGGGTACACAGCAATGCCCTACTCTAGTCCAGTGTAGTCTGGGCTCATACAA 1718
 Db
 80 CTGGGTGACCAAGCAATGCCCTACTCCAGTCCCTCCATGTCTGGGCTTCCCAA 26

Search completed: December 15, 2004, 11:16:14
Job time : 204.055 secs

; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21090
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21090

Query Match 9.8%; Score 242.4; DB 4; Length 245;
Best Local Similarity 99.6%; Pred. No. 1.6e-55;
Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 551 CTACTTTCCTCTTACAGAACTTTGGAAACCAAGGAGCAGACCTTTAGCTGCTGTT 610
DB 1 CTACTTTCCTCTTACAGAACTTTGGAAACCAAGGAGCAGACCTTTAGCTGCTGTT 60

QY 611 TTATGCTATTTGACCAAGGCTACATATCTCGGTCAGTGTGATCTTTGATAATGAAG 670
DB 61 TTATGCTATTTGACCAAGGCTACATATCTCGGTCAGTGTGATCTTTGATAATGAAG 120

QY 671 GCATTGCTATTTTGCACTTCAGTTCACATCTATTTATGGTAAATCTGTAATAAATG 730
DB 121 GCATTGCTATTTTGCACTTCAGTTCACATCTATTTATGGTAAATCTGTAATAAATG 180

QY 731 GGTCAATTTTGGACAAATGCTGCTGCTTATCTATATGCTCTGCTGCTGCTGCTG 790
DB 181 GGTCAATTTTGGACAAATGCTGCTGCTTATCTATATGCTCTGCTGCTGCTGCTG 240

QY 791 GTGG 794
DB 241 GTGG 244

RESULT 13

US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12856
; LENGTH: 914
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-12856

Query Match 9.6%; Score 237.4; DB 4; Length 914;
Best Local Similarity 64.2%; Pred. No. 7.9e-54;
Matches 374; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 1741 TTAGATGATTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACCA 1800
DB 2 TTCGATGACTTCCGCGAGGCGCTACTGCTGCAGATGAACACTCCGAGGACGCTCGC 61

QY 1801 GTAATGCTCTGGTGGATTTAGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTTG 1860
DB 62 ATAATGCTCTGGTGGGACTACGGCTACAGATGAACGCAATCGGCAATCGGAGATTA 121

QY 1861 GTGGATATAACACTGGAATAACAGGCAATAGCACTGTGGGAAAGCTATGCTTCT 1920
DB 122 GTGGATAACAATCTTGAACAAACACATATATCGCGCGCTGGCCAGCGATGGCCTCT 181

QY 1921 AATGAACAGCAGGCTATATAAATCATGAGGACTCTAGATGATATTATGTTTGGTTAT 1980
DB 182 TCGGAGGAGAGAGCTACGAGATATCGGGAAGCTGATGCTGACCTGCTTCTGAT 241

QY 1981 TTGGAGGGTTATGGCTATCTGGTGATGATATCAAAATTTCTCTGATGGTTAG 2040

DB 242 TTCGAGGGCTCACTGGCTACTCATCGGACGATATCAACAGTTCTCTGTGGATGGTGGC 301
QY 2041 ATAGCTGAAGGAGAAATCCCAAG---ACATTGGGAAAGTGAATTTTACCCACAG 2097
DB 302 ATTGGCGGACGACGGATCGTGGTGGCAGATCCGCGAAAGGACTACTATCGGCGCAAC 361
QY 2098 GGAGAAATTCGCTGTAGACAAAGCAGATCCCTACTTTGTTGAATTCCTTATGATATAA 2157
DB 362 GGAGAGTTCCGAGTGGACAAAGAGGGCTCACCCACATGCTCAATTTGTTGATACAG 421
QY 2158 ATGTCACTACTACAGATTTGGAGAAATGCGAGCTGGATTTTCGTACACCCCGAGTTTTCAC 2217
DB 422 ATGTGCTACTATCGCTTTGGGCAATGTACACGAGGTGGCAAGCCCGAGGCTACCAT 481
QY 2218 CGAACAGCTAATGCTGAGATTGGAATTAAGGACATTAATTCARACATTTGGAAGAAGCC 2277
DB 482 CGAGTTCTGCTCCCGCAGATCGGCAACAAAGGACTTTGAACTGGATGCTCTCGGAGGGCG 541
QY 2278 TTTACATCAGAACACTGGCTTGTAGGATATATAAGTAAAG 2320
DB 542 TACACCAAGGAGCACTGGCTGGTGGCACTACACAGTTAAAG 584

RESULT 14

US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-1318

Query Match 8.1%; Score 201.4; DB 4; Length 560;
Best Local Similarity 63.9%; Pred. No. 2.9e-44;
Matches 342; Conservative 0; Mismatches 181; Indels 12; Gaps 2;

QY 1196 TTATTGCATCAGTGTCTGAGCATCACTACACCTGAGTGGTGTCTTCTTTGATCTAC 1255
DB 560 TCATTGATCCGTCGGAGCATGCCACACTTGGTTCTGTTCTTCTTTGATCTGC 501

QY 1256 ATATTCTTGTATGATCTTCCCAGAGGCTTTGGTTCTGCTCAATAATATCAACGATG 1315
DB 500 ACATCATGTTGTGGCTTCCCAGTGGAGTGTGGTACTGTCATCAAGCAGATCAACGACG 441

QY 1316 AAGAGATTTGTTGCTCTATATGCAATCAGTGTCTGCTACTTCTGCTGAGTGGTGC 1375
DB 440 AGCGGTTTTCGTTGGTCTGTACGCCATCAGTGGGTTTACTTCGCTGTGTGATGGTGC 381

QY 1376 GACTGATGTTGACTTTTCACTCCAGTCTGTGTATGCTGTCTCAATTCCTTTTCAAAATG 1435
DB 380 GTTGTGATTTGACCTCAGCCCGTGTGTGATGCTGCGCGAGTGGGCTTTTCGGAC 321

QY 1436 TTTTGGACACTATTGGGGGATGA-----CATGAAAGGAAATCCACCTGTGGAG 1488
DB 320 TGTGATGTTTCTCGAAGAGGATTCGTCTAAGCAATGGGAGCAGCATCAACGCGAG 261

QY 1489 -----GACAGCAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAG 1543
DB 260 CCACCGAAGTGGATGAAGCTGAGGATTCATTTGAGAAGAGAGCGCTGTACGCAAGGCTG 201

QY 1544 GTAAAGTGAAGAAACATGCACTGACAGGAAAAAACTGAAGAGGATTAGGCCCTAATA 1603
DB 200 GCAAGCTGAAGCATCTGCTACTAAGCATGATGCCCGCAGGAGTACTGGCTGCTCAACC 141

RESULT 10

US-09-248-796A-3089
; Sequence 3089, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089

; LENGTH: 867

; TYPE: DNA

; ORGANISM: Candida albicans

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (25)

; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

Query Match 11.0%; Score 272.2; DB 4; Length 867;

Best Local Similarity 63.5%; Pred. No. 3.2e-63;

Matches 434; Conservative 0; Mismatches 243; Indels 6; Gaps 1;

QY 1648 TTGCTGTCCTACTGCTGGTCCACGAAGCAAGCTCTACTAGTCCAAAGTGTAGTCCTG 1707

DB 85 TTGTTTACATGTTGCTGGTAAACATCGAATGCTTATCATCAACATCAGTTGTTTA 144

QY 1708 GCCTCATCAATCAATGATGGCACCGAGAAATATCTTAGATGATTTTGAAGCTTACTTT 1767

DB 145 GCATCCGAAACCCAGATGGCTCACACATATCATTTGATGATATATAGAAAGCTATTAC 204

QY 1768 TGGCTTAGGCAAAATACAGATGAACATGACGAGTAAATGCTTGGGGGATTTAGCTAT 1827

DB 205 TGGTTAGAATGAATACACGAAGATCCCAAGTTATGGCTGTTGGGATTTATGTTAT 264

QY 1828 CAGATAGTGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGAATAACAGC 1887

DB 265 CAAATCGGGGTATGGCTGTAGAACCCACACTTGTGTATAACAAATACATGGAATAACACA 324

QY 1888 CACATAGCACTGGTGGGAAAGCTATGCTCTTAATGAACACAGCAGCCTATAAAATCATG 1947

DB 325 CATATGCCACTGTTGTAAGGCNAUGTCTTCCCTGGAAGATGTGCTGATGAAATTTG 384

QY 1948 AGGACTCPAGATGATGATATGTTTGGTTATTTTGGAGGGGTTATTCGCTATTCGT 2007

DB 385 AGACAACACGATGTTGATTTAGTTATATTTTGGAGGGTATTGGGTTTATTCGT 444

QY 2008 GATGATATCAAAATTTCTGATGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 2067

DB 445 GATGATATCAAAATTTCTGATGTTAGGATAGCTGAAGGAGAACATCCCAAGAC 504

QY 2068 ATTCGGGAACTGACTATTTTACCCCAACAGGAGAAATTCGTTAGACAAAGAGATCC 2127

DB 505 ATCAAGAAAGAGACTACTTTTACTGACCGAGGAGAAATAAAGTGGATAAAGATGCATCA 564

QY 2128 CCTACTTTGTTGATGCTTATGTAATAAATGTCTATACATACAGATTTGGAGAAATGCAG 2187

DB 555 CTGGCAATGAAGAAATCTTTGATGATTAAGTTATCGTAATCATAGATCTGAA----- 618

QY 2188 CTGATTTTCTGACACCCCGAGTTTGAACGACAGCTAATGCTGAGATGGAATGAAG 2247

DB 619 TTGTTGGAGGTAGAGATGTTGTTGATAGAGTTAGAAACCAACAAATCCAGCCATGAA 678

QY 2248 GACATTAATTCAAACATTTTGGAGAGAGCTTTTACATCAGAACACTGGCTGTTAGGATA 2307

Db 679 GTACCGAAATGAAATGTTGTTCAAGAGCGCTTCACATCAGAAAATTTGATTGTGAGAATT 738

QY 2308 TATAAAGTAAAGACCTTGATTA 2330

Db 739 TACAAGTTAAAGATTTGGATTA 761

RESULT 11

US-09-513-999C-1438

; Sequence 1438, Application US/09513999C

; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59:US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 1438

; LENGTH: 250

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 54..248

US-09-513-999C-1438

Query Match 10.1%; Score 250; DB 4; Length 250;

Best Local Similarity 100.0%; Pred. No. 1.4e-57;

Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 794 GTTATGTTTATCATCAATCTTATCCACTGCATGATTTTGTGTTTACTGATGCGA 853

DB 1 GTTATGTTTATCATCAATCTTATCCACTGCATGATTTTGTGTTTACTGATGCGA 60

QY 854 GATACAGCAAGAGCTTACATAGCATATAGCACTTCTACATGCTGGTTTATATAT 913

DB 61 GATACAGCAAGAGAGCTTACATAGCATATAGCACTTCTACATGCTGGTTTATATAT 120

QY 914 CAATGAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTG 973

DB 121 CAATGAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGCGAGCTG 180

QY 974 CAGGTCTCTTGCATGCTGCAAGCTTATGCTTCTTCAGTATCTGAGAGACCGATTA 1033

DB 181 CAGGTCTCTTGCATGCTGCAAGCTTATGCTTCTTCAGTATCTGAGAGACCGATTA 240

QY 1034 CAAAACAAGA 1043

DB 241 CAAAACAAGA 250

RESULT 12

US-09-513-999C-21090

; Sequence 21090, Application US/09513999C

; Patent No. 6783961

GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59:US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

QY 1089 TGTGTTCTTCTAGTGTCTATTTTGGTCTATATACAGG 1124
 DB 1346 GGTGTTCTGTTGGTACCTCTGCTCAGGCTTACCGG 1381

RESULT 8
 US-09-270-767-158/c
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 158
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Query Match 11.0%; Score 273.8; DB 4; Length 1660;
 Best Local Similarity 63.8%; Pred. No. 1.8e-63;
 Matches 432; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
 QY 1798 CGAGTAATGCTCTGTTGGTGGATTTGGCTATCAGATAGCTGGAATGGCTAATAGAACTACG 1857
 DB 829 CCAGTAGTAAGCCTCTCTGAAATCGTCTAAATGTTGCGGAATGCAACAGACGACG 770
 QY 1858 TTGTTGGATATACACCTGGAATTAACAGCCATAGCAGTCTGCTGGTGGAAAGCTATGCT 1917
 DB 769 CTAGTGGATAATAATAGCTGGAACAATAGTCAATAGCCTGCTGGCAAGGCAATGCT 710
 QY 1918 TCTAATGAACAGCAGCCTCTAAATATCATGAGGACTCTAGATAGATTATGTTTGGTT 1977
 DB 709 TCAACCGAGGAGAGTCTCTCGAAATATGACATCTCTTCACTGGACTACGTTTTGGTG 650
 QY 1978 ATTTTGGAGGGTTATGGCTATCTCTGATGATATCAACAAATTTCTCTGGATGTT 2037
 DB 649 ATCTTTGGCGGTGATCGGCTATTCGGCGATGATATCAACAAATTTCTCTGGATGTT 590
 QY 2038 AGGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAG 2097
 DB 589 CGAATTCGGAGGAGAGCATCCCAAGACATTAAGGAAGCGATTACTTTACCGACCG 530
 QY 2098 GGAGAAATTCGCTGACAAAGCAGGATCCCTACTTTGTAATTCGCTTATGTATAA 2157
 DB 529 GGTGAATTCAGGTAGATGCCAGAGTCTCCGGCCCTGCTCACTGCTTATGTACAA 470
 QY 2158 ATGTCTACTACAGATTGGAGAAATGCAGTCTGATTTTGGTACACCCCGAGTTTGAC 2217
 DB 469 TTAAGCTACTACAGATTGGGAAATTTGAAGTTGGACTACAGAGGCCATCTGGATATGAT 410
 QY 2218 CGACAGTATGCTGAGATTGGAAATAGGACATTAATTCAAATTTGGAGAGCC 2277
 DB 409 CGCACAGTAAACCGCTCATTTGGGAATAGGACTTTGATCTGCTACTCTGGAGAGGCC 350
 QY 2278 TTTACATCAGAACTGGCTTTGTTAGGATATATAAGTAAAGCACTGATACAGGGAG 2337
 DB 349 TACACCACAGAACTGGCTTTGTCATCTATAGTGTGAAGAGCGGATGTTCAAT 290
 QY 2338 ACATTAGATCACAACCTCGAGTCAACCAATTTCCCAAAACAGAGTATTTGTCAGAG 2397
 DB 289 AGACCATCAGTGAAGACAAGAGAGAGAACGATTCCTCCAGCA ---AATCTATCTCGAGA 233
 QY 2398 AAGACTACCAAAAGGAGCGTGGCTACATTTAAAAATAAGCTGGTTTTTAAGAAAGGCAAG 2457
 DB 232 AAGAACTCGAAGCGTCGCAAGGGCTACATACGAAACCGACCGGTTGTTTGAAGGAAAA 173
 QY 2458 AAAATACTCAGAGAC 2474

DB 172 CGAACCTTGAATAAATAC 156

RESULT 9
 US-09-270-767-15440/c
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

Query Match 11.0%; Score 273.8; DB 4; Length 1660;
 Best Local Similarity 63.8%; Pred. No. 1.8e-63;
 Matches 432; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
 QY 1798 CGAGTAATGCTCTGTTGGTGGATTTGGCTATCAGATAGCTGGAATGGCTAATAGAACTACG 1857
 DB 829 CCAGTAGTAAGCCTCTCTGAAATCGTCTAAATGTTGCGGAATGCAACAGACGACG 770
 QY 1858 TTGTTGGATATTAACACCTGGAATTAACAGCCATAGCAGTCTGCTGGTGGAAAGCTATGCT 1917
 DB 769 CTAGTGGATAATAATAGCTGGAACAATAGTCAATAGCCTGCTGGCAAGGCAATGCT 710
 QY 1918 TCTAATGAACAGCAGCCTCTAAATATCATGAGGACTCTAGATAGATTATGTTTGGTT 1977
 DB 709 TCAACCGAGGAGAGTCTCTCGAAATATGACATCTCTTCACTGGACTACGTTTTGGTG 650
 QY 1978 ATTTTGGAGGGTTATGGCTATCTCTGATGATATCAACAAATTTCTCTGGATGTT 2037
 DB 649 ATCTTTGGCGGTGATCGGCTATTCGGCGATGATATCAACAAATTTCTCTGGATGTT 590
 QY 2038 AGGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAG 2097
 DB 589 CGAATTCGGAGGAGAGCATCCCAAGACATTAAGGAAGCGATTACTTTACCGACCG 530
 QY 2098 GGAGAAATTCGCTGACAAAGCAGGATCCCTACTTTGTAATTCGCTTATGTATAA 2157
 DB 529 GGTGAATTCAGGTAGATGCCAGAGTCTCCGGCCCTGCTCACTGCTTATGTACAA 470
 QY 2158 ATGTCTACTACAGATTGGAGAAATGCAGTCTGATTTTGGTACACCCCGAGTTTGAC 2217
 DB 469 TTAAGCTACTACAGATTGGGAAATTTGAAGTTGGACTACAGAGGCCATCTGGATATGAT 410
 QY 2218 CGAACAGTAACTCTGAGATTGGAAATAGGACATTAATTCAAATTTGGAGAGGCC 2277
 DB 409 CGCACAGTAAACCGCTCATTTGGGAATAGGACTTTGATCTGCTACTCTGGAGAGGCC 350
 QY 2278 TTTACATCAGAACTGGCTTTGTTAGGATATATAAGTAAAGCACTGATACAGGGAG 2337
 DB 349 TACACCACAGAACTGGCTTTGTCATCTATAGTGTGAAGAGCGGATGTTCAAT 290
 QY 2338 ACATTAGATCACAACCTCGAGTCAACCAATTTCCCAAAACAGAGTATTTGTCAGAG 2397
 DB 289 AGACCATCAGTGAAGACAAGAGAGAGAACGATTCCTCCAGCA ---AATCTATCTCGAGA 233
 QY 2398 AAGACTACCAAAAGGAGCGTGGCTACATTTAAAAATAAGCTGGTTTTTAAGAAAGGCAAG 2457
 DB 232 AAGAACTCGAAGCGTCGCAAGGGCTACATACGAAACCGACCGGTTGTTTGAAGGAAAA 173
 QY 2458 AAAATACTCAGAGAC 2474
 DB 172 CGAACCTTGAATAAATAC 156

; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Query Match 13.7%; Score 340.2; DB 4; Length 1660;

| | Best Local Similarity | 67.08; | Pred No. 2.2e-81; | Mismatches | Indels | Gaps | 1; |
|----|-----------------------|------------------------|---------------------------|-------------------------------|--------|------|----|
| | Matches | 499; | Conservative | 0; | | | |
| QY | 1730 | CCAGGAATATCTTAGATGATT | TAGAGAAGCTTACTTTTGCTAAGGC | AAATAACAGATG | 1789 | | |
| Db | 789 | CCCGCAACATTTTAGACGATTT | CAGAGAGCTTACTACTGGCTTT | CGCAGAACACTCCGC | 848 | | |
| QY | 1790 | AACATGCACGAGTAATCTCT | TGTGGTGGGATTATGGCTATC | CAGATAGCTGGGATGCTAATA | 1849 | | |
| Db | 849 | ATGATGCTCGGCTTATGTCT | TGTGGTGGATTACGATACC | GAATAGCGGGAATGCCAACA | 908 | | |
| QY | 1850 | GAACATACGTTGGTCGGATAA | CAACACCTGGAAATACAGCC | CACATAGCACTGGTGGGAAAAG | 1909 | | |
| Db | 909 | GAACAGCCTAGTGGATAATA | TACGTGGAAACAATAGTC | CACATAGCGCTGGTGGCAAGG | 968 | | |
| QY | 1910 | CTATGCTCTTCAANTGAAC | AGCAGCCTATATAAATCAT | GAGGACTCTAGATGTGAATTATG | 1969 | | |
| Db | 969 | CAATGCTCTTCAACCGAG | GAGAGTGCTTACGAATAT | TGACATCTCTTGACGTGGACTACG | 1028 | | |
| QY | 1970 | TTTTGGTATATTTTGGAGGG | GTATTGGCTATCTCGTGAT | GATATCAACAAATTTCTCT | 2029 | | |
| Db | 1029 | TTTTGGTGATCTTTGGCG | GTGTGATCGGCTATTTCT | GGCGATGATCAACAAGTTCTCTG | 1088 | | |
| QY | 2030 | GSATGGTTAGGATAGCTGA | AGGAGAGACATCCCAA | GACATTCGGGAAAGTGACTATTTTA | 2089 | | |
| Db | 1089 | GGATGGTCCGAAATTTGCT | GAGGAGAGCATCCC | NAAGGACATTAAGGAAAGCGAATTA | 1148 | | |
| QY | 2090 | CCCCACAGGGAGAAATCC | CGTGTAGACAAAGCAG | ATCCCTACTTTTGTGAATTTGCCTTA | 2149 | | |
| Db | 1149 | CGACCGCGGTGAATTC | GAGGTAGATGCCGAGGTGCT | CCGGCCCTGCTCAACTGCCTTA | 1208 | | |
| QY | 2150 | TGATATAAATGTCTATCT | ATACAGATTTGGAGAAAT | GCAGCTGGATTTTCGTACACCCCCAG | 2209 | | |
| Db | 1209 | TGTAACAAATTAAGCTACT | ATCAGATTTGGGGAAAT | TGAAGTTGGACTACAGAGGTCCATCTG | 1268 | | |
| QY | 2210 | GTTTTGGCCGAAACAGTAA | TCGTGAGATTTGGAAATA | TAAGGACATTAATATTC | 2269 | | |
| Db | 1269 | GATATGATCGCACACGCT | TAACGCGTATTTGGGAA | TAAGGACATTCGATCTGACCTACTCG | 1328 | | |
| QY | 2270 | AAGAAGCCTTTATCATC | AGAACACATGGCTTTGT | TAGGATATATAAGTATAAAGCACCTGATA | 2329 | | |
| Db | 1329 | AGGAGGCCCTACACCA | CAGAACACTGGCTGTT | CGCATCTATAGGTTGAAGAGCCGCATG | 1388 | | |
| QY | 2330 | ACAGGGAGACATTAGATCA | CAACACCTCGAGTCAC | CCAACATTTTCCCAAAACAGAGATATT | 2389 | | |
| Db | 1389 | AGTTCAATAGACCATCTA | CTGAGACCAAGGAGAGA | ACGAATTCCTCCAGCAACTTCATTT | 1448 | | |
| QY | 2390 | TGTCAAAGAGACTACCA | AAAGGAAGCGTGGCT | TACATTAANAATAANGCTGTTTTTAAGA | 2449 | | |
| Db | 1449 | CGAGAAAGAACTCTA | AGGCTCGGAAG-- | -GGCTACATACGAAACCGCGGTCTTGTTA | 1505 | | |
| QY | 2450 | AGGCGAAGAAATATCT | TAAGAGAC | 2474 | | | |
| Db | 1506 | AGGGAAACCGAACCTT | GAAATAAAC | 1530 | | | |

RESULT 7

```

US-09-270-767-11648
; Sequence 11648, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1998-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11648

```

| | | |
|-----|---|-----|
| 113 | AGTGGCTGGGTACAGCGCCTAATCACTTCCGCACTCTGCTTAATCGCCTGGCTGGCGG | 172 |
| 251 | GCTTCAGCTCGGCTCTTGGCCGTATCGCGTTCGAAAGCATCATCCACGAGTTTCGACC | 310 |
| 173 | GATTTTCCTCTCGGCTCTTGGCCGTATCGCGTTCGAGTCGATTATCCATGAGTTGATC | 232 |
| 311 | CGTGGTTAACTATAGATCAACACATCATCTTCGCACTCTCATGGTTCATGAATTTTAA | 370 |
| 233 | CGTGGTTCAACTACCGGGCCHCCGCCCTCATGGTGCAGAAATGGTTCGTTCAACTCCCTCA | 292 |
| 371 | ATTGGTTTGATGAAGAGCATGGTATCCACTAGAAAGATPAGTAGTGCTACTGTTTACC | 430 |
| 293 | ACTGGTTCGACGCGCGCATGGTATCGGCTCGGAGGATTTGGGCGGTACCGTCTATC | 352 |
| 431 | CAGGGTTGATGATTAACCGCTGGCCCTTATTCATTGGGATTTTAAATACATTGAACATAACTG | 490 |
| 353 | CGGCGCTGATGATTACGTCGGCGGGAATCCATTGGCTGCTGCGACTCAACATACCGG | 412 |
| 491 | TTCACTAAGAGAGGTATGTGTGTTCTCTTGACCAACTTTTAGCGGCCCTTACATCTATAT | 550 |
| 413 | TCCATATTCTGATGACATCTGCGGTGTTCTTGCGGCGGATCTTCAGTGGCCTGCACTCCA | 472 |
| 551 | CTACTTTCCTGCTTAAAGAGAACTTTTGGAAACCAAGAGCAGGACTTTTAGCTGCTTGT | 610 |
| 473 | CCACCTACTCTGTCGACCAAGAGAGCTGTGGTCCGCGGCGCGGCCCTCTTTCCGCCGCA | 532 |
| 611 | TTATTGCTATTGTACCAAGCTPACATATCTCGGTTCAGTACGTGATCCTTTGATTAATGAAG | 670 |
| 533 | TCATCGCCATCGTGGCTGGCTACATCAGTAGGTGCGTGGCTGGATCGTAGGATACGAGG | 592 |
| 671 | GCATTGCTATTTTGCATCTCAGTTCACATACTATTATGGGTAAAACTCTGTAAAACTG | 730 |
| 593 | GCATTGCCATATTGGCCCTGCGGTTCCACTACTTCTGTGGGTGGCTCAGTGAAGACTG | 652 |
| 731 | GGTCAGTTTTTGGACAAATGTGCTGCTATCTTATCTATATGCTCTCTGCTGGG | 790 |
| 653 | GATCGGTGTTCTGGTGGCGCGACCGGTTTGTCTTACTTCTACATGGTGTCCGCGCTGG | 712 |
| 791 | GTGGTTATGTATTTATCATCAATCTTATCCACTGCATGTATTGTGGTTGTACTGATGC | 850 |
| 713 | GTGGCTACGTGTTTCATCATCAACCTGATACCCCTGCACGCTCTTCGTACTGCTCATATGG | 772 |
| 851 | AGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTGTGGGGTTTAAT | 910 |
| 773 | GCAGGTACTCGCCCGCTCTGCTGACCGAGCTACAGCACTTCTACATCTCTGGGAGCTGT | 832 |
| 911 | TATCAATCGAGATACCTTTTGTGGGATTCAGCGCAATCGAAACAAAGTGAACATCGGAG | 970 |
| 833 | TCTCCATCGAGATCCCTTCTGTGGGATTTCCAAACCGATACGCCACCTGTAACACATGGCTG | 892 |
| 971 | C | 971 |
| 893 | C | 893 |

RESULT 5

US-09-270-767-158
: Sequence 158 Application US/09270767

; sequence 138, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION

APPLICANT: Homburger

; TITLE OF INVENTION: Nuclear

FILE REFERENCE: 7326-094

FILE REFERENCE: 7328-094
CURRENT APPLICATION NUMBER: US/09/270 767

;; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17

17-03-1999 FILING DATE: 1999-03-17
NUMBER OF SEC ID NOS: 0000000000

; NUMBER OF SEQ ID NOS: 62517

```

; SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 158

; LENGTH: 1660

TYPE: DNA

ORGANISM: *Drosophila melanogaster*

US-09-270-767-158

Query Match 13.7%: Score 340.2: DB 4: Length 1660:

Query Match 13.7%; Score 340.2; DB 4; Length 1000;
Best Local Similarity 67.0%; Pred. No. 2.2e-81;

BEST LOCAL SIMILARITY 67.0%; PRED. NO. 2.2E-81;
Matches 499; Conservative 0; Mismatches 243; Index 3;

| | Matches | 499; | Conservative | 0; | Mismatches | 243; | Indels | 3; | Gaps | 1; |
|----|---------|-------------------------|---|------|------------|------|--------|----|------|----|
| QY | 1730 | CCAGGAATATCTTAGATGATTTT | TAGAGAAGCTTACTTTGCTGAAGCAAAATACAGATG | 1789 | | | | | | |
| DB | 789 | CCCGCAACATTTTAGACGATTT | CAGAGAGGCTTACTACTGGCTTTCAGAAACACTGCCG | 848 | | | | | | |
| QY | 1790 | AACATGCACAGATAATGTCT | TCGTCGGATATGGCTATCAGATAGCTGGAAATCGCTAATA | 1849 | | | | | | |
| DB | 849 | ATGATGTCGCGTATGCTCT | TGTCGGATACCGATACCAGATAGCGGAAATGCGCAACA | 908 | | | | | | |
| QY | 1850 | GAATAGCTTTGGTGGATAATA | CACCTGGAAATACAGCCACATAGCACTCGTGGGAAAG | 1909 | | | | | | |
| DB | 909 | GAACGACGCTAGTGGATAATA | TACGTGGAAACAATAGTCATAGCCCTGGTTGGCAAG | 968 | | | | | | |
| QY | 1910 | CTATGCTCTTAATGAACAG | CAGCAGCTATAAAATCATGAGACTCTAGATGTAGATATATG | 1969 | | | | | | |
| DB | 969 | CAATGCTCTCAACCGAGG | AGAAGTCCTACGAAATATATGACATCTCTTGACGTGGACTACG | 1028 | | | | | | |
| QY | 1970 | TTTTGGTATTTTGGAGGGT | TATTGGCTATTTCTGGTATGATATCAACAARAATTTCTCT | 2029 | | | | | | |
| DB | 1029 | TTTTGGTGAATTTTGGCGGT | GTGATCGGCTATTTCTGGGATGATATCAACAAGTTCCTGT | 1088 | | | | | | |
| QY | 2030 | GGATGTTTAGGATAGCTGA | AGGAGAACAATCCCAAAGACATTCGGGAAATGCACTATTTTA | 2089 | | | | | | |
| DB | 1089 | GGATGTTCCGAAATTGCT | GAGGAGACATCCCAAGCAATTAAGGAAAGCGATTACTTTTA | 1148 | | | | | | |
| QY | 2090 | CCCCACAGGGAGAAATTC | CGTGTAGACAAAGCAGAGATCCCTACTTTCTGTCAAATGGCTTA | 2149 | | | | | | |
| DB | 1149 | CCGACCCCGTGAATTCAG | GTATGATCGGAAGTGTCTCCGGCCCTGCTCAACTGCTTA | 1208 | | | | | | |
| QY | 2150 | TGTTATAAATGTGCATCTA | CTACAGATTTGGAGAAATCGAGCTGGATTTTCGTACACCCCCAG | 2209 | | | | | | |
| DB | 1209 | TGTACAAATTAAGCTACT | ACAGATTCGGGGAATTTGAAGTTGGACTACAGAGGTCCATCTG | 1268 | | | | | | |
| QY | 2210 | GTTTTTGACCGAACACGCT | TAATGCTGAGATTTGGAATTAAGGACATTAATTCAAACATTTGG | 2269 | | | | | | |
| DB | 1269 | GATATGATCCACACCTAC | CGCCGTCATTTGGGAATAAGGACTTCGATCTGCACCTACCTGG | 1328 | | | | | | |
| QY | 2270 | AAGAAGCCTTTACATCAGA | ACACTGCTTTGTAGGATATATAAGTAAAAAGCACTTGATA | 2329 | | | | | | |
| DB | 1329 | AGGAGGCCTACACCCAGA | ACACTGCTTTGTTCGATCTATAGGGTGAAGAAGCCGCGATG | 1388 | | | | | | |
| QY | 2330 | ACAGGGAGACATTAGATCA | CAAAACCTCGAGTCCACCAACATTTTCCCAAAACAGAAGTATT | 2389 | | | | | | |
| DB | 1389 | AGTTCAATAGACCATCACT | GAAGACCAAGGAGAGAACGATTTCTCCAGCAAACTTCATTT | 1448 | | | | | | |
| QY | 2390 | TGTCAAAGAGACTACAAAG | AGGACGTGGCTACATTAAATAAGCTCGTTTTTAAGA | 2449 | | | | | | |
| DB | 1449 | CGAGAAAGAATCTTAAG | CGTCCGAAG---GGCTACATACGAAACCGGACCCGGTTGTTGTTA | 1505 | | | | | | |
| QY | 2450 | AAGCGAAGAAATATCTA | AGAAGAC | 2474 | | | | | | |
| DB | 1506 | AGGGAACCGAACCTTGA | ATAAAC | 1530 | | | | | | |

RESULT. T. 6

RESULT 6
US-09-270-767-15440

US-09-270-767-13440 : Sequence 15440. Application US/09270767

Sequence 15440, Appl. No. 6703491

Patent No. 6703491
GENERAL INFORMATION.

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and pro

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/0

CURRENT FILING DATE: 1999-03-

: NUMBER OF SEO ID NOS: 62517

! NUMBER OF SEQUENCES: 1000000
: SOFTWARE: Patern

; SOFTWARE: PAC

```

; SEQ ID NO 15
; LENGTH: 16

```

; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Query Match 20.4%; Score 507; DB 4; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.2e-126;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1683 CTACTCTAGTCCCAAGTCTAGTCTCTGGCTCATACATCATGATGGCACCAGGAATATCTT 1742
DB 1 CTACTCTAGTCCCAAGTCTAGTCTCTGGCTCATACATCATGATGGCACCAGGAATATCTT 60
QY 1743 AGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAAGT 1802
DB 61 AGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAAGT 120
QY 1803 AATGCTTGGTGGGATATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGCT 1862
DB 121 AATGCTTGGTGGGATATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGCT 180
QY 1863 GGATAATAACACCTGGGAATTAACAGCCACATAGACACTGGTGGGAAAGCTATCTCTTAA 1922
DB 181 GGATAATAACACCTGGGAATTAACAGCCACATAGACACTGGTGGGAAAGCTATCTCTTAA 240
QY 1923 TGAACAGCAGCCTATATAATCATGAGGCTCTAGATGATGATATGTTTGGTTATTT 1982
DB 241 TGAACAGCAGCCTATATAATCATGAGGCTCTAGATGATGATATGTTTGGTTATTT 300
QY 1983 TGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTGGATGGTTAGAT 2042
DB 301 TGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTGGATGGTTAGAT 360
QY 2043 AGCTGAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGAGA 2102
DB 361 AGCTGAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGAGA 420
QY 2103 ATTCCGTTGAGCAAAAGCAGGATCCCTACTTTTGGTGAATTCGCTTATGATATAAAATGTC 2162
DB 421 ATTCCGTTGAGCAAAAGCAGGATCCCTACTTTTGGTGAATTCGCTTATGATATAAAATGTC 480
QY 2163 ATACTACAGATTGGGAAATGCAGCT 2189
DB 481 ATACTACAGATTGGGAAATGCAGCT 507

RESULT 3
US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Forz, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08

; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(616)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

Query Match 16.7%; Score 415.4; DB 3; Length 616;
Best Local Similarity 95.8%; Pred. No. 6.5e-102;
Matches 436; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
QY 622 GTACCAAGGTACATATCTCGGTCTAGTCTGATGATCTTTGATATGAAGGCAATGCTATT 681
DB 5 GTACCAAGGTACATATCTCGGTCTAGTCTGATGATCTTTGATATGAAGGCAATGCTATT 64
QY 682 TTTCACATTCAGTTCACATATCTTTATGGGTAAATCTGTAAAACTGGGTCAAGTTTTT 741
DB 65 TTTCACATTCAGTTCACATATCTTTATGGGTAAATCTGTAAAACTGGGTCAAGTTTTT 124
QY 742 TGAACAATGTGCTGCTTATCTTATATGTTCTCTGTTGGGTGGTTATGTA 801
DB 125 TGAACAATGTGCTGCTTATCTTATATGTTCTCTGTTGGGTGGTTATGTA 184
QY 802 TTTCATCATCAATCTTATTCCTGCTGATGATCTTTGTTGTTTACTGATGACAGATACAGC 861
DB 185 TTTCATCATCAATCTTATTCCTGCTGATGATCTTTGTTGTTTACTGATGACAGATACAGC 244
QY 862 AAAAGAGTCTCATAGCATAATAGCACTTTTACATGTTGGGTTTAAATATTAATCAATGCG 921
DB 245 AAAAGAGTCTCATAGCATAATAGCACTTTTACATGTTGGGTTTAAATATTAATCAATGCG 304
QY 922 ATACCTTTTGTGGGATTCAGCAATCAAGCAAGTGAACATGAGGAGCTGCAAGTGC 981
DB 305 ATACCTTTTGTGGGATTCAGCAATCAAGCAAGTGAACATGAGGAGCTGCAAGTGC 364
QY 982 TTTCATTCCTGCAAGCTTATGCTTTTTCGAGTATCTGAGAGCCGATTAACAAACAA 1041
DB 365 TTTCATTCCTGCAAGCTTATGCTTTTTCGAGTATCTGAGAGCCGATTAACAAACAA 423
QY 1042 GAGTTCACACCTTTTCTTTTGGGTTGATCACT 1076
DB 424 GAGTTCACACCTTTTCTTTTGGGTTGATCACT 458

RESULT 4
US-09-270-767-12331
; Sequence 12331, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12331
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12331

Query Match 14.6%; Score 361.8; DB 4; Length 900;
Best Local Similarity 66.5%; Pred. No. 2.3e-87;
Matches 519; Conservative 0; Mismatches 262; Indels 0; Gaps 0;
QY 191 AGCCGGCTGGTGGGAGTCTCTCTCTTCAACCATCTCTCTCTGGCTGGTGGTGGCG 250

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:40:24 ; Search time 198.055 Seconds
(without alignments)
8903.949 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgccccc.....tatctaagaagactgttttaa 2481

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCT_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 539.4 | 21.7 | 2157 | 4 | US-09-614-221A-318 |
| 2 | 507 | 20.4 | 507 | 4 | US-09-513-999C-1965 |
| 3 | 415.4 | 16.7 | 616 | 3 | US-09-328-111-332 |
| 4 | 361.8 | 14.6 | 900 | 4 | US-09-270-767-12331 |
| 5 | 340.2 | 13.7 | 1660 | 4 | US-09-270-767-158 |
| 6 | 340.2 | 13.7 | 1660 | 4 | US-09-270-767-15440 |
| 7 | 309.6 | 12.5 | 1386 | 4 | US-09-270-767-11648 |
| 8 | 273.8 | 11.0 | 1660 | 4 | US-09-270-767-15440 |
| 9 | 273.8 | 11.0 | 1660 | 4 | US-09-270-767-15440 |
| 10 | 272.2 | 11.0 | 867 | 4 | US-09-248-796A-3069 |
| 11 | 250 | 10.1 | 250 | 4 | US-09-513-999C-1438 |
| 12 | 242.4 | 9.8 | 245 | 4 | US-09-513-999C-21090 |
| 13 | 237.4 | 9.6 | 914 | 4 | US-09-270-767-12856 |
| 14 | 201.4 | 8.1 | 560 | 4 | US-09-270-767-1318 |
| 15 | 201.4 | 8.1 | 560 | 4 | US-09-270-767-16600 |
| 16 | 176.6 | 7.1 | 594 | 4 | US-09-248-796A-3090 |
| 17 | 173.8 | 7.0 | 503 | 4 | US-09-270-767-27262 |
| 18 | 149.4 | 6.0 | 487 | 4 | US-09-702-705-1655 |
| 19 | 149.4 | 6.0 | 487 | 4 | US-09-736-457-1655 |
| 20 | 149.4 | 6.0 | 487 | 4 | US-09-614-124B-1655 |
| 21 | 149.4 | 6.0 | 487 | 4 | US-09-671-325-1655 |
| 22 | 149.4 | 6.0 | 487 | 4 | US-09-658-824-1655 |
| 23 | 147 | 5.9 | 307 | 4 | US-09-702-705-375 |
| 24 | 147 | 5.9 | 307 | 4 | US-09-702-705-1271 |
| 25 | 147 | 5.9 | 307 | 4 | US-09-736-457-375 |
| 26 | 147 | 5.9 | 307 | 4 | US-09-736-457-1271 |
| 27 | 147 | 5.9 | 307 | 4 | US-09-614-124B-375 |

| | | | | | | |
|----|-------|-----|-----|---|---------------------|--------------------|
| 28 | 147 | 5.9 | 307 | 4 | US-09-614-124B-1271 | Sequence 1271, Ap |
| 29 | 147 | 5.9 | 307 | 4 | US-09-671-325-375 | Sequence 375, App |
| 30 | 147 | 5.9 | 307 | 4 | US-09-671-325-1271 | Sequence 1271, App |
| 31 | 147 | 5.9 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 32 | 147 | 5.9 | 307 | 4 | US-09-658-824-375 | Sequence 375, App |
| 33 | 147 | 5.9 | 307 | 4 | US-09-658-824-1271 | Sequence 1271, Ap |
| 34 | 145.4 | 5.9 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| 35 | 145.4 | 5.9 | 307 | 4 | US-09-736-457-588 | Sequence 588, App |
| 36 | 145.4 | 5.9 | 307 | 4 | US-09-614-124B-588 | Sequence 588, App |
| 37 | 145.4 | 5.9 | 307 | 4 | US-09-671-325-588 | Sequence 588, App |
| 38 | 145.4 | 5.9 | 307 | 4 | US-09-589-184-588 | Sequence 588, App |
| 39 | 145.4 | 5.9 | 307 | 4 | US-09-658-824-588 | Sequence 588, App |
| 40 | 137.4 | 5.5 | 487 | 3 | US-09-385-982-213 | Sequence 213, App |
| 41 | 134 | 5.4 | 292 | 4 | US-09-313-294A-4442 | Sequence 4442, Ap |
| 42 | 123 | 5.0 | 291 | 4 | US-09-313-294A-4834 | Sequence 4834, Ap |
| 43 | 104.6 | 4.2 | 288 | 4 | US-09-313-294A-6705 | Sequence 6705, Ap |
| 44 | 100.8 | 4.1 | 302 | 4 | US-09-313-294A-6869 | Sequence 6869, Ap |
| 45 | 96 | 3.9 | 414 | 4 | US-09-513-999C-271 | Sequence 271, App |

ALIGNMENTS

RESULT 1

US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614, 221A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

| | | | | | | |
|-----------------------|-------|--|--------------|------------|--------|--------------------|
| Query Match | 21.7% | Score | 539.4 | DB 4 | Length | 2157 |
| Best Local Similarity | 55.1% | Pred | No. 5.3e-135 | | | |
| Matches | 1237 | Conservative | 0 | Mismatches | 896 | Indels 110; Gaps 5 |
| Qy | 175 | TCGGGGGGCTGTCGACGCCGCTGGTGGCAGTGGTTCCTCTCCATCACCCTCTTC | 234 | | | |
| Db | 7 | TCGACCGGTGCTGTGTTTGTCTGTGTTTTCAGACCATCTCTCAAGCTCGTCTTCGTG | 66 | | | |
| Qy | 235 | CTGGCTGGCTGGCGGCTTCAGCTCGGCTCTTTCGGCGTCCGCTTCGAAAGATC | 294 | | | |
| Db | 67 | CGATTTTGGGGCTGCCATATCATCAGTTTGTTCAGTCATCAATTTGATCTATT | 126 | | | |
| Qy | 295 | ATCCAGAGTTCGACCCGCTGGTAACTATAGATCAACACATCATCTTGATCTCATGGG | 354 | | | |
| Db | 127 | ATCCATGAATTCGACCCCTGTTCAATTTATAGGGTACCAATATCTCGTCAACAATTCG | 186 | | | |
| Qy | 355 | TTCTATGAATTTTAAATTTGTTGATCAAAAGAGCATGTATCCACTAGGAAGATAGTA | 414 | | | |
| Db | 187 | TTTACAGTTTGTGACTGTTTGACGCGTACCTGTTACCCCTCGGAGGTTACT | 246 | | | |
| Qy | 415 | GGTGTACTGTTTACCCAGGGTGTGATGATAACCGCTGGCCCTTATT---CATGGAATTTA | 471 | | | |
| Db | 247 | GGAGGACTTTTATATCTCTGGTTGATGACGACTAGTGGTTCATCTGGCAGCCCTCGC | 306 | | | |
| Qy | 472 | AATACATTCAACATAACTGTTTCACATAGAGACGATGTGTGTTTCCCTTGGACCACTTTT | 531 | | | |
| Db | 307 | AACTGGTGGGTTGCCCATTTGACATCAGAACGTTTGTGTGCTATTGTGGCCACTATT | 366 | | | |

QY 2156 AAATGTCATACAGATTTCGAGAAATGCGAGCTGGATTTTCGTACACCCCGAGTTTG 2215
 Db 2091 AATTAAGCTACTACAGATTTCGGGAATGAAGTTGGACTACAGAGTCCATCTGGATATG 2150
 QY 2216 ACCGAACAGTAATCTCGATGTTGGAATAGGACATTAATCAACATTTTGGAGAGAG 2275
 Db 2151 ATCGCACAGTAAACCGCTCATTTGGGAATAGGATTCGATCTGACCTACCTGGAGAGG 2210
 QY 2276 CTTTACATCAGAACACATCTGGCTTTGTTAGGATATATAAAGTAAAGACACCTGTATAACAGGG 2335
 Db 2211 CCTACACACAGAACATCTGGCTTTGTTAGGATATATAAAGTAAAGACACCTGTATAACAGGG 2335
 QY 2336 AGACATTAGTACACAAACCTCGAGTACCAACATTTTCCCAAAACAGAGATTTTGTCAA 2395
 Db 2271 ATAGACCATCACTGAGACCAAGAGAGAGACATTTCTCCAGCAACTTCATTTCCAGAA 2330
 QY 2396 AGACATCTACCAAAAGGAGCGTGGCTACATATAAATAAGCTGGTTTTTAAGAAAGGCA 2455
 Db 2331 AGACTCTAAGCGTCGCAAG--GGCTACATACGAAACCGACCGGTTGTTGTTAAGGGAA 2387
 QY 2456 AGAAATATCTAAGAGAC 2474
 Db 2388 AACGAACCTTGAATAAAC 2406

RESULT 14

AX136135 2510 bp DNA linear PAT 30-MAY-2001
 DEFINITION Sequence 57 from Patent EP1067182.
 ACCESSION AX136135
 VERSION AX136135.1 GI:14272543

SOURCE Homo sapiens (human)
 ORGANISM

REFERENCE
 AUTHORS Ota,T., Isegai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.
 TITLE Secretory protein or membrane protein
 JOURNAL Patent: EP 1067182-A 57 10-JAN-2001;
 Helix Research Institute (JP)

FEATURES
 source
 1..2510
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 94..954
 /note="unnamed protein product"

CDS

/codon_start=1
 /protein_id="CAC39710.1"
 /translation="MLMLMLMPFAVHCTWVTNAYSSPSVILASYNHDTNRILDDF
 REAYFLQRNTDEHARVMSWDYGYQAGVANRTTLVDNNWNHSHIALVGMASNE
 TAAFKIMRPLDLYVLVTFGGVIGSGDINKFLMVRVIAEGHPKIDIESYFTPOG
 EPRVDKAGSPILLNCLMYKMSYRFGEMQLDFRTPGDFDRNAEIGNDKIDFKILEE
 APTSEHLVRLVYKVKAPDNRETLDKPRVTNIPPKQKYLKSKTKRGVINKKLVEK
 KKKKISKKT"
 ORIGIN

Query Match 38.5%; Score 954; DB 6; Length 2510;
 Best Local Similarity 100.0%; Pred. No. 2e-178;
 Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1528 TTGTATGATAGGCGAGGTAAAGTGGAGGAAACATGCAACTCAACAGGAAAAAACTGAAGAG 1587
 Db 1 TTGTATGATAGGCGAGGTAAAGTGGAGGAAACATGCAACTCAACAGGAAAAAACTGAAGAG 60
 QY 1588 GGATTTAGGCCCTTAATATAAAGAGATTGTCCACCATGTTGATGCTGATGCTATTGATGATG 1647
 Db 61 GGATTTAGGCCCTTAATATAAAGAGATTGTCCACCATGTTGATGCTGATGCTATTGATGATG 120

QY 1648 TTTGCTGTCCACHTGACCTGGGTCAAGAGCAATGCTTACTTCTAGTCAAGTGTAGTCTGT 1707
 Db 121 TTTGCTGTCCACHTGACCTGGGTCAAGAGCAATGCTTACTTCTAGTCAAGTGTAGTCTGT 180
 QY 1708 GCCTCATACAATCATGATGGCACACAGGAATATCTTAGATGATTTTAGAGAGCTTACTTT 1767
 Db 181 GCCTCATACAATCATGATGGCACACAGGAATATCTTAGATGATTTTAGAGAGCTTACTTT 240
 QY 1768 TGCTTAAGGCAAAATACAGATGAACATGACGAGTATGTTCTGTTGGGATTTAGGCTAT 1827
 Db 241 TGCTTAAGGCAAAATACAGATGAACATGACGAGTATGTTCTGTTGGGATTTAGGCTAT 300
 QY 1828 CAGATAGCTGGAATGGCTAATAGAACTAGTGTGGGATAATAACACCTGGAAATAACAGC 1887
 Db 301 CAGATAGCTGGAATGGCTAATAGAACTAGTGTGGGATAATAACACCTGGAAATAACAGC 360
 QY 1888 CAGATAGCTGTTGGGAAAGCTATGTTCTTANTGAACAGCAGCCTATAAATCATG 1947
 Db 361 CAGATAGCTGTTGGGAAAGCTATGTTCTTANTGAACAGCAGCCTATAAATCATG 420
 QY 1948 AGGACTCTAGATGATGATTTGTTGTTGTTGTTGAGGGGTTATTTGGCTATTCTGGT 2007
 Db 421 AGGACTCTAGATGATGATTTGTTGTTGTTGTTGAGGGGTTATTTGGCTATTCTGGT 480
 QY 2008 GATGATATCAACAAATTTCTCTGGATGTTAGTAGCTGAAGGAGAACATCCCAAAGAC 2067
 Db 481 GATGATATCAACAAATTTCTCTGGATGTTAGTAGCTGAAGGAGAACATCCCAAAGAC 540
 QY 2068 ATTCCGGAAAGTGACTATTTTACCCACAGGAGAAATTCCTGTAGACAAAGCAGGATCC 2127
 Db 541 ATTCCGGAAAGTGACTATTTTACCCACAGGAGAAATTCCTGTAGACAAAGCAGGATCC 600
 QY 2128 CCTACTTTGTTGAATGCTTATGATATAAATGTCATACTACAGATTTGGAGAAATGCAG 2187
 Db 601 CCTACTTTGTTGAATGCTTATGATATAAATGTCATACTACAGATTTGGAGAAATGCAG 660
 QY 2188 CTGATTTGCTACACCCCGAGTTTGAACGAAACACGTATGCTGAGATTTGGAATTAAG 2247
 Db 661 CTGATTTGCTACACCCCGAGTTTGAACGAAACACGTATGCTGAGATTTGGAATTAAG 720
 QY 2248 GACATTAATTCAAACATTTGGAAGAGGCTTTTACATCAGAACACTGCTGTTGTTAGGATA 2307
 Db 721 GACATTAATTCAAACATTTGGAAGAGGCTTTTACATCAGAACACTGCTGTTGTTAGGATA 780
 QY 2308 TATAAGTAAAGCACCTGATTAACAGGAGACATTTAGATCAACAACTCGAGTCAACAC 2367
 Db 781 TATAAGTAAAGCACCTGATTAACAGGAGACATTTAGATCAACAACTCGAGTCAACAC 840
 QY 2368 ATTTTCCCAAAACAGAGATTTTGTCAAGAGAGACTACCAAAAGGAGAGCGTGGCTACAT 2427
 Db 841 ATTTTCCCAAAACAGAGATTTTGTCAAGAGAGACTACCAAAAGGAGAGCGTGGCTACAT 900
 QY 2428 AAAAAAGCTGTTTTTAAAGCAAGAAAATATCTTAGAGACACTGTTTAA 2481
 Db 901 AAAAAAGCTGTTTTTAAAGCAAGAAAATATCTTAGAGACACTGTTTAA 954

RESULT 15
 B0123520 2510 bp DNA linear PAT 18-SEP-2002
 LOCUS B0123520
 DEFINITION Secretory protein or membrane protein.
 ACCESSION B0123520
 VERSION B0123520.1 GI:23218465
 KEYWORDS JP 2002017376-A/29.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2510)
 AUTHORS Ota,T., Isegai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.
 TITLE Secretory protein or membrane protein
 JOURNAL Patent: JP 2002017376-A 29 22-JAN-2002;

Db 462 CCACCTACTCTGCTACCAAGAGCTGTGGTCCGGGCGCGGCTCTTGGCGCGAGT 521
Qy 611 TTATTGCTATTGTACCAAGCTACATATCTCGGTGAGTGTGGATCCTTTGATTAAG 670
Db 522 TCATCGCATCTGCTGCTGCTACATCAGTAGTGGTGGTGGATCGTAGCAACGAG 581
Qy 671 GCATTGCTATTGTTGCACTTCAGTTCATACATATTTATGGTAAATCTGTAAGACTG 730
Db 582 GCATTGCTATTGTTGCACTTCAGTTCATACATATTTATGGTAAATCTGTAAGACTG 641
Qy 731 GGTGAGTTTGTGGCAATGCTGCTGCTTATCCTATTCTATATGCTGCTGCTGG 790
Db 642 GATCGGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
Qy 791 GTGGTATGTTATATCAATCAATCTTATTCACCTGCTGCTGCTGCTGCTGCTGCTG 850
Db 702 GTGGTATGTTATATCAATCAATCTTATTCACCTGCTGCTGCTGCTGCTGCTGCTG 761
Qy 851 AGAGTACAGCAAAAGAGTCTACATAGATATAGCACTTTCTCATGTTGGTAAATAT 910
Db 762 GCAGGTACTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 821
Qy 911 TATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAAAGTGAACATGCGCAG 970
Db 822 TCTCCATGAGATCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 881
Qy 971 CTGAGGTGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1030
Db 882 CGCTGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 941
Qy 1031 TAACAAAACAGAGTTCAGACCTTTCTTTTGGGCTGATCACTAGCTGCAAGTGGCTG 1090
Db 942 TGTGCGCAACGAGTTCGGAAGCTGTTTCATCGTGGCGGCTGCTGCTGCTGCTGCTG 1001
Qy 1091 TCTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
Db 1002 TCTTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1061
Qy 1151 TTTATTCTATGTTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1210
Db 1062 TCTACTGCTGTTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1121
Qy 1211 CTGAGCATCAACCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1270
Db 1122 CGAGCATCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1181
Qy 1271 CCTTCCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1330
Db 1182 CTTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241
Qy 1331 CTCTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1390
Db 1242 TCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
Qy 1391 TGACTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1450
Db 1302 TCACGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1361
Qy 1451 TGGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1498
Db 1362 TGCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1421
Qy 1499 ATGAGGATGACAAAAGAACCAAGAAATTTGATGATGAGGAGGCTGAGGAGGAAAC 1558
Db 1422 AAGCTGAGGATTTCCATTGAGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1481
Qy 1559 ATGCAACTCAACAGGAAAGAACTGAAGAGGATGAGGCTGCTGCTGCTGCTGCTGCTGCTG 1618
Db 1482 GTACTAAGCATGATCCACAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1541
Qy 1619 CCATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1678

Db 1542 TTTTGGCGGTCTTATGCTGTTGATGATGTTGCTGCTGCCACTGCACGTGGGTGACAGCA 1601
Qy 1679 ATGCTACTACTAGTCCAAAGTGTAGTCTTGGCTCATACAA---TCATGATGGCCACAGGA 1735
Db 1602 ATGCTACTACTAGTCCAAAGTGTAGTCTTGGCTCATACAAAGTGTAGTGGATCCCGCA 1661
Qy 1736 ATATCTAGATGATTTAGAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATG 1795
Db 1662 ACATTTAGACGATTCAGAGAGGCTTACTACTGCTTCGACAGACACTGCGGATGATG 1721
Qy 1796 CACGAGTAATGCTTGGTGGGATTTAGCTATCAGATAGCTGGAATGGCTAATAGAACTA 1855
Db 1722 CTCGCTGATGCTTGGTGGGATTTAGCGATACAGATAGCGGATGCGCAACAGACAGCA 1781
Qy 1856 CGTTGTTGATATACACCTCGAATTAACAGCCACATAGCACTGCTGGGAAAGCTATGT 1915
Db 1782 CGCTAGTGGATTAATAGCTGGAACATAGTCAATAGCCCTGTTGGCAGGCAATGT 1841
Qy 1916 CTTCTAATGAACAGAGCTTATAAATCATGAGGACTCTAGATGTAGATTTATTTTGG 1975
Db 1842 CTTCAACCGAGGAGAGTCTTACGAAATTAAGCATCTCTTGGCTGGACTACGCTTTGG 1901
Qy 1976 TTATTTTGGAGGGTATTGCTTATTTGGTGTATGATATCAACAAATTTCTCTGATGG 2035
Db 1502 TGAATTTGGCGGTGATCGGCTATTTCTGGGATGATATCAACAGTCTCTGTGGATGG 1961
Qy 2036 TTAGGATAGTGAAGAGAGCAATCCAAAGACATTCGGGAAAGTGAATTTTACCCAC 2095
Db 1962 TCGAAATTTCTGAGGAGAGCATCCAAAGGACATTAAGGAAAGGATTTACTTTACCGACC 2021
Qy 2096 AGGAGAAATTCGCTGTAGACAAAGCAGGATCCCTTACTTTTGAATTTGCTTATGTATA 2155
Db 2022 GCGTGAATTCAGGTTAGATGCCGAAGTGTCTCGGCTGCTCACTGCTTATGTACA 2081
Qy 2156 AATGTCATACTACAGATTTGGAAGATGCACTGCTGATTTTCTGACACCCCGAGTTTGG 2215
Db 2082 AATTAAGTACTACAGATTTGGAAGATGCACTGCTGATTTTCTGAGGTTCCATCTGGATG 2141
Qy 2216 ACCGAACAGTATGCTGAGATTTGGAATTAAGGACATTAATTTCAAACATTTGGAAGA 2275
Db 2142 ATGCAACAGTAAACGCGCTCAITGGGAATAGGACTTCGATCTGCTGCTGAGGAGG 2201
Qy 2276 CTTTACATCAGACACTGCTGCTGATGATATATAAGGATTAAGGACCTGATACAGGG 2335
Db 2202 CTTACACCAAGACACTGCTGCTGATTTGCACTTAGGTTGAAGAGCCGATGATGCA 2261
Qy 2336 AGACATTAGATCACAACCTCGAGTCAACCAATTTCCCAAAACAGAGATTTTGTCAA 2395
Db 2262 ATAGACATCACTGAAGACCAAGAGAGAGAACGATTCCTCCAGCAACTTCATTTCCAGAA 2321
Qy 2396 AGAGACTACCAAG 2455
Db 2322 AGAATCTTAGCGTCCAGAG---GGCTATACAGAAACCGGACCGGTTGTTTAAAGGAA 2378
Qy 2456 AGAAATATCTAAGAGAG 2474
Db 2379 AAGCAACCTTGAATAAAC 2397

RESULT 13

CQ589353
LOCUS CQ589353 2699 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 17111 from Patent WO0171042.

ACCESSION CQ589353

VERSION CQ589353.1 GI:41648215

KEYWORDS

SOURCE

ORGANISM

Prosophila sp.

Prosophila sp.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Phylogeny; Prosophilidae; Prosophila.

1

REFERENCE

AUTHORS

Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.

1062 TCTACTCGCTGGGATACCTGGCTACGCCAAGATCCACATTCCTCCATTCATTCGATCCGCTG 1121
QY 1211 CTGACATCAACCTAGACCTGGGCTCTCTCTCTTGTATCTACATATTCCTGATGTA 1270
Db 1122 CGGACATCAGCCACCACTGGCTCTCTCTCTTGTATCTGCACTCTGCTGGTGG 1181
QY 1271 CTTCCAGAGAGGCTTGGCTCTGATCAAAATATCAACGATGAAGAGTATTGTTG 1330
Db 1182 CTTCCAGAGGAGTGGTGGTCTGATCAAGCAGATCAACGAGAGCGCTTTTCGTTG 1241
QY 1331 CTCTATATGCAATCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1390
Db 1242 TCTCTAGCCATCAGTGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1301
QY 1391 TGACTCCAGTCTGCTGATGCTCTGCAATTCGCTTTTCAATGTTTTTGAGCACTATT 1450
Db 1302 TCACGCCGCTGGTGTGATGCTGGCGAGTGGCTTTTCGGGACTGTTGGATGTTGCC 1361
QY 1451 TGGGGGATGA-----CATGAAAGGGAATCCACCTCTGGAG-----GACAGCAGTG 1498
Db 1362 TGCAAGAGGATCTCTAAGCGATGGCACAGCCATAGCGCAGCCAGTGGATG 1421
QY 1499 ATGAGATGCAAAAGAAACCAAGAAATTTGTATAGAGGAGTAAAGTGAGGAAC 1558
Db 1422 AAGCTGAGGATTCATTTAGAGAGAGCGCTGTACGACAGAGCTGGCAAGCTGAAGCATC 1481
QY 1559 ATGCAACTGAACAGAGAAACTCAAGAGGATAGGCCCTTAATAAAAAAGCATTTGCA 1618
Db 1482 GTACTAAGCATGATGCCAGCAGATCTGGCTGAGTCCACCTGAAGATTTGTTA 1541
QY 1619 CATGTTGATGCTGATGCTATGATGTTTGTGTCACCTGCTGCTGCTGCTGCTGCTGCT 1678
Db 1542 TTTTGGCGCTTCTAATGCTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601
QY 1679 ATGCTACTCTAGTCTCAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1735
Db 1602 ATGCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1661
QY 1736 ATATCTTAGATGATTTAGAGAACTTACTTTTGGCTAGGCAATACAGATGACATG 1795
Db 1662 ACATTTAGAGATTTAGAGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1721
QY 1796 CACGAGTAACTCTGCTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTA 1855
Db 1722 CTGCGTTATGCTTGTGGGATTTAGGATACAGATAGCGGATAGCGGATAGCGGATAGCG 1781
QY 1856 CGTTGGTGAATTAACACCTGGAATTAACAGCACATAGCCTGCTGGGAAAGCTATGT 1915
Db 1782 CGCTAGTGAATTAACAGTGGAACTAGTACATAGGCTGCTGGTGGCAAGCAAGTGT 1841
QY 1916 CTTCTAATGAACAGCAGCTTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGG 1975
Db 1842 CTTCAACCGAGAGAGTCTTACGAAATTTAGATCTCTGAGTGGACTAGCTTTGG 1901
QY 1976 TTATTTTGGAGGCTTATGGCTATCTGCTGATGATACAAATTTCTCTGATGG 2035
Db 1902 TGATCTTTGGCGGTGATCGGCTATTTCTGGCATGATTAACAAGTTCCTGATGG 1961
QY 2036 TPAGATAGCTGAAGAGAAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCAC 2095
Db 1962 TCCGAATTTCTGAGGAGAGCATCCCAAGACATTAAGGAAAGCGATTAATTTTACCGACC 2021
QY 2096 AGGAGATTCGGTGTAGCAAAACAGGATCCCTACTTTGTTGATTCCTCTTATGATA 2155
Db 2022 GCGGTGAATTCAGGTAGATGCCAAGTGTCTCGGCCCTGCTCACTGCTCTTATGATA 2081
QY 2156 AAATGCTACTATACAGATTTGGAGAAATGAGTGGATTTTCTGACACCCAGGTTTG 2215
Db 2082 AATTAAGTACTATACAGATTTGGGAAATTAAGTTGGAATACAGAGGTCCATCTGATATG 2141
QY 2216 ACCGAACAGTAACTGCTGAGATTTGGAATAAGGACATTAATTAACAACTTTGGAGAG 2275
Db 2142 ATCGCACACGTAACGCCGCTCATTGGGAATTAAGGACTTCGATCTGACTCTGCTGAGGAG 2201

2276 CTTTACATCAAGAACACATGGCTTGTAGGATATATAAAGTAAAGACACCTGTATAACAGG 2335
Db 2202 CCTACACACAGAACACATGGCTTGTTCGATCTATAGGTTGAAGAGCCGATGAGTTCA 2261
QY 2336 AGCATTAGATCAAAACTCGAGTCAACCAATTTTCCAAACACAGAGTATTGTTCAA 2395
Db 2262 ATAGACCATCACTAGAGAACCAAGGAGAGAGCATTTCTCCAGCAAACTTCATTTCAGAA 2321
QY 2396 AGAAGACTACCAAAAGAGGAGGCTGGCTACATTAATAAAGCTGGTTTTTAAGAAAGGCA 2455
Db 2322 AGAATCTTAAGCGTGCAG--GCTACATACGAAACCGACCGTGTGTTGTTAAGGAA 2378
QY 2456 AGAATATCTTAAGAGAC 2474
Db 2379 AAGCACTTGAATAAAC 2397

RESULT 12
AX799088
LOCUS 2417 bp DNA linear PAT 08-OCT-2003
DEFINITION Sequence 7 from Patent WO03054008.
ACCESSION AX799088
VERSION AX799088.1 GI:37605061
KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 Perreault, C. and McBride, K.
Mammalian SIMP protein, Gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 7 03-JUL-2003;
Competitive Inc. (CA)
Location/Qualifiers
1..2417
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/note="STT3, Acc# AF132552"

JOURNAL
FEATURES
source
Query Match 38.5%; Score 955.8; DB 6; Length 2417;
Best Local Similarity 64.8%; Pred. No. 8.7e-179;
Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;

ORIGIN
QY 191 AGCGGCTGGTGGGAGTGGCTTCTCTCTCACATCTCTCTCTCTCTCTCTCTCTCTCTCT 250
Db 102 AGGTGGCTGGCTACAGAGCCCTAATCACTTCGCCATCCCTGCTAACTGCTGGCTGGCG 161
QY 251 GCTTCAGCTCGCGCTCTTCGCCCTTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACC 310
Db 162 GATTTTCT 221
QY 311 CTTGTTTAACTATAGATCAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
Db 222 CTTGTTTAACTATAGATCAACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 281
QY 371 ATTGTTTGAAGAGAGATGGTATCCATCAGGAGAAATAGTAGTGGTGGTGGTGGTGGTGG 430
Db 282 ACTGTTTGAAGAGAGATGGTATCCGCTCGCAGGATTTGGGCGGTACCGTCTATC 341
QY 431 CAGGTTGATGATAACCGCTGGCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 490
Db 342 CCGGCTGATGATTAAGTTCGCGGGAATCCATTTGGTCTGCTGCTGCTGCTGCTGCTGCTG 401
QY 491 TTCACATAGAGAGATGATGTTGTTCTTGTGCAACCACTTTTAGCGGCTTTTACATCTAT 550
Db 402 TCCATATTCGTGACATCTCGGTGTTCTCTGCGCGGCTCTTCTGTTGCTGCTGCTGCTGCT 461
QY 551 CTACTTTCTGCTTACAGAGAACTTTTGGAAACCAAGGAGGACTTTTAGCTGCTGTTGT 610

AF132552.1 GI:4689327
FLI CDNA.
Drosophila melanogaster (fruit fly)
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1. (bases 1 to 2417)
Tsang, G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M.,
Lewis, S.E., Suh, C. and Rubin, G.M.
Direct Submission
Submitted (02-MAR-1999) Berkeley Drosophila Genome Project,
University of California Berkeley, 539 Life Sciences Addition,
#3200, Berkeley, CA 94720, USA
REFERENCE
2. (bases 1 to 2417)
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J.,
Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G.,
Miranda, A., Mungall, C.J., Nunco, J., Pacleb, J., Paragas, V., Park, S.,
Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M.
and Celisner, S.
Direct Submission
Submitted (12-FEB-2003) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(<http://fruitfly.berkeley.edu>) or send email to
cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1. 2417
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
1. 2417
/gene="OstStt3"
/db_xref="FLYBASE:FBgn0011336"
1. 2395
/gene="OstStt3"
/note="Longest ORF"
/codon_start=2
/product="Gm01839p"
/protein_id="A027851.2"
/db_xref="GI:28317200"
/translators="LSREVAFQIGNFVETGWNMDNRTPKMNSKVAGYSSLLTF
ALLIAMLGFSRLFAVIFESI IHEFPWENYRATAYVONGWVNFINKPDERAWY
PLGRVGTVPGLMTSGGHLHLLVHNPVHIDICVFLAIPESGLTSYLLTK
ELWAGAGLAFIAIPGVYISRSVAGSYDNEGAI PAQTYFLVRSVKTLSVFW
SAAALSYFVAVSANGGVYINLIPLHVFLLTMGRYSFRLTYSITFYILGLLFSM
QIPVGPQPIRTSEHMAUGVFLVLMVATLHRLQSVLSRNEFRKLFIVGGLLVGV
FVAVVLTGLVAVPSMGREYSLMDTGYAKIHPIIASVSEHPTWTFSPFDLHLY
CAPFGVYICIKINDEPVFLVLYA:SAVYFAGVMVRLMLTLPVVMAGVAFSGLL
DVFLQEDSCKRMGTIASAEVDEAEDSIEKTLVDKAGLKHRTKHDQADQTVSSN
LKSIVILAVMLMFAVHCTVWTSNAYSPSVILAFHNSODGSNILLDDREAYVYL
SONTADBARVMSWDYGYQIAGMANETTLVDNWNNSHIALVAGMAMSTBEKSEVM
TSLVDVYVLI FGVYIGSDDI NKFVWVIAEGHPKDIKESYFFIDRGEPRVDAE
GAPALLNCLMYKLSYRFGLKLDYRGPGRDTRNAVIGNKDFDLTYLEAYITTEH

ORIGIN
Query Match 38.5%; Score 955.8; DB 3; Length 2417;
Best Local Similarity 64.8%; Pred. No. 8.7e-179;
Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
QY 191 AGCGCGTGGTGGCAGTGGCTTCTCTCTTCCATCATCTCTTGGCATCTCATGGGTCTTATGAATTTTAA 370
DB 102 AGTGGTGGCTACAGCAGCCTAATCACCTTGGCCATCTGCTAATCGCTGGCTGGCGG 161
QY 251 GCTTCACTCGCGCTTTCGCGCTCATCGCTTGGAAAGCATATCCACAGAGTTCGAC 310
DB 162 GATTTTCTCTCGCTTTCGCGCTCATCGCTTGGATCGATTAATCATGATGTTGATC 221
QY 311 CGTGGTTTAACTATAGATCAACATCATCTTCTGCACTCTCATGGGTCTTATGAATTTTAA 370
DB 222 CGTGGTTTCACTACCGCGGCGACCGCTCATAGTTCGAGAATGGTGGTACAACTTCCTCA 281
QY 371 ATTGGTTTGTATGAAGAGCATGTATCCACTAGGAGAAATAGTAGTGTGTTACTGTTTACC 430
DB 282 ACTGGTTCGAGCAGCGCATGTATCCGCTCGGAGAGTTGTGGCGGTACCGTCTATC 341
QY 431 CAGGTTGATGATTAACCGCTGGCTTATTCATTTGATTTTAAATACATTTGAACAATCTG 490
DB 342 CCGGCTGATGATTAACCGCTGGCTTATTCATTTGATTTTAAATACATTTGAACAATCTG 401
QY 491 TCCACATTAAGAGAGCATGTATGTGTTCTTCTTGGCACCACCTTTTAGCGGCTTACATCTAT 550
DB 402 TCCATATTCTGACATCTGCGTGTCTTCTGGCGCGCATCTTCACTGAGTGTGAGTCTCATCT 461
QY 551 CTACTTTCTGCTTCAAGAGAACTTTTGGAAACAAGGAGCAGGACTTTTAGTGTGTTT 610
DB 462 CCACCTACTGCTGACCAAGAGAGCTGTGCTCGCGCGCGCGCTTCTTGGCGCGCAGCT 521
QY 611 TTATTGCTATTGTACAGGCTCATATCTCGCTGCTAGTGTGATCTCTTGGTATGATGAG 670
DB 522 TCATCGCATCGTCTGCTACATCAGTAGTGTGCTGGTGGTGGATCGTACGATACGAGG 581
QY 671 GCATTGCTATTGTCACCTTCACTGATCATATATTTATGGTAAATCTGTAAATACTG 730
DB 582 GCATTGCTATTGTCACCTTCACTGATCATATATTTATGGTAAATCTGTAAATACTG 641
QY 731 GGTCAAGTCTTTTGGCAATGTGCTGCTGTATCTTCTATTTCTATATGCTCTCTGCTGG 790
DB 642 GATCCGTTCTGCTCGCGCGCAGCGCTTGTCTCTACTTCTACATGTTGCTCGCTGG 701
QY 791 GTGGTTATGTTATTCATCAATCTTATTCACCTGATGTATTTGTTGTTGTTTACTGATGC 850
DB 702 GTGGCTACGTGTTTCATCAACCTGATACCCCTGCAGCTCTTCTGCTACTCTCATTTAG 761
QY 851 AGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGTTGGTTTAAAT 910
DB 762 GCAGGTACTCGCGCGCTGCTGTCACCGCATCAGCACCTTCTACATCTTGGAGCTGCTGT 821
QY 911 TATCAATGAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGGCAG 970
DB 822 TCTCCATGAGATACCTTTTGTGGGATTCAGCCCAATCAGAACAAAGTGAACACATGGCAG 881
QY 971 CTGCAAGTGTCTTGTCAATGCTGCAAGTGTATGCTTTCTTGGCATGTCTGAGAGACCGAT 1030
DB 882 CGCTGGAGTGTGTTGTGCTCTTATGGCGGTGGCCACCTTGGGCCATTTGCACTCGCTGC 941
QY 1031 TAACAAAAACAAGAGTTCAGACCCCTTTCTTTTGGGTGTATCTACTAGCTCGAGGTGCTG 1090
DB 942 TGTCCGCGCAACAGAGTTCGGAAGCTGTTTCATCGTCGGGGAGTTGCTGTGGCGCTGGCG 1001
QY 1091 TGTTCCTTAGTGTCACTATTGACTTTATACAGGTTTACATTCACCATGGAGTGGAGTGGCAGGT 1150
DB 1002 TCTTTTGGCGCTCGTGTGCTCACAATGCTGGGCTTGTGGCCCGCTGGAGTGGAGCCT 1061
QY 1151 TTTATTCAATTGGGGATATCGGATATGCAAAATAACATTCCTCAATTTATTCATGATGT 1210

QY 2050 GGAGAACATCCCAAGACATTCGGAAAGTGCATTTTACCCACAGGAGAAATCCGT 2109
 DB 541 GGAGAACATCCCAAGACATTCGGAAAGTGCATTTTACCCACAGGAGAAATCCGT 600
 QY 2110 GTAGCAAAAGCAGGATCCCTACCTTGTGAAATGGCTTATGTATATAAATGTCTACTAC 2169
 DB 601 GTAGCAAAAGCAGGATCCCTACCTTGTGAAATGGCTTATGTATATAAATGTCTACTAC 660
 QY 2170 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 2229
 DB 661 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 720
 QY 2230 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGAGCCTTTTACACAGAA 2289
 DB 721 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGAGCCTTTTACACAGAA 780
 QY 2290 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATTAACAGGAGACATTAATACAC 2349
 DB 781 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATTAACAGGAGACATTAATACAC 840
 QY 2350 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 2409
 DB 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 900
 QY 2410 AGGAGCTGTCTACATTAATAAGTAAAGTGGTTTAAAGAGGCAAGAAATATCTAG 2469
 DB 901 AGGAGCTGTCTACATTAATAAGTAAAGTGGTTTAAAGAGGCAAGAAATATCTAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972

RESULT 10

BD063986 2546 bp DNA linear PAT 27-AUG-2002
 LOCUS Secreted proteins and polynucleotides encoding them.
 DEFINITION BD063986
 ACCESSION BD063986
 VERSION BD063986.1 GI:22609589
 KEYWORDS JP 2001506848-A/10.
 SOURCE unidentified
 ORGANISM unclassified.

REFERENCE 1 (bases 1 to 2546)
 AUTHORS Jacobs,K., McCoy,J.M., Lavallie,B.R., Racie,L.A., Verberg,D.,
 Treacy,M., Spaulding,V. and Agostino,M.J.
 TITLE Secreted proteins and polynucleotides encoding them
 JOURNAL Patent: JP 2001506848-A 10 29-MAY-2001;
 GENETICS INSTITUTE INC
 COMMENT PN JP 2001506848-A/10
 PD 29-MAY-2001
 PF 12-DEC-1997 JP 1998525986
 PR 13-DEC-1996 US 08/766263
 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
 DAVID VERBERG,
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
 C12N15/12,C12N5/10,C07K14/47,A61K38/17
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key Location/Qualifiers.

FEATURES
 source
 1..2546
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

ORIGIN
 Query Match 39.2%; Score 972; DB 6; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 5.4e-182;
 Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1510 AAAAGAAACCAAGGAAATTTGTATCATAAAGGACGTTAAAGTGAAGGAAACATGCAACTGAA 1569

DB 1 AAAAGAAACCAAGGAAATTTGTATCATAAAGGACGTTAAAGTGAAGAAACATGCAACTGAA 60
 QY 1570 CAGGAAAAAATCTGAGAGGGATTAGCCCTTAATATAAAAGCANTGTCCACATGTTGATG 1629
 DB 61 CAGGAAAAAATCTGAGAGGGATTAGCCCTTAATATAAAAGCANTGTCCACATGTTGATG 120
 QY 1630 CTGATGCTATTCGATGATGTTTCTCTCCACTGTACTCTGGGTCAACAAGCAATGCTACTCT 1689
 DB 121 CTGATGCTATTCGATGATGTTTCTCTCCACTGTACTCTGGGTCAACAAGCAATGCTACTCT 180
 QY 1690 AGTCCAAAGTGTAGTCTGCTCATCAATCATGATGTCACAGCAATCTTAGATGAT 1749
 DB 181 AGTCCAAAGTGTAGTCTGCTCATCAATCATGATGTCACAGCAATCTTAGATGAT 240
 QY 1750 TTTAGAGAAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAAATGCTCT 1809
 DB 241 TTTAGAGAAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAAATGCTCT 300
 QY 1810 TGTGGATTTGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTGTTGGTGAAT 1869
 DB 301 TGTGGATTTGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTGTTGGTGAAT 360
 QY 1870 AACACCTGGAATAACAGCCACATAGCACTGGGTGGGAAAGCTATGCTTCTTAATGAAACA 1929
 DB 361 AACACCTGGAATAACAGCCACATAGCACTGGGTGGGAAAGCTATGCTTCTTAATGAAACA 420
 QY 1930 GCAGCCTTAATAATCATGAGGACTCTAGATGATGATTTGTTGGTATTTTGGAGGG 1989
 DB 421 GCAGCCTTAATAATCATGAGGACTCTAGATGATGATTTGTTGGTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTTCTGGTATGATTAACAATAATTTCTCTGATGGTTAGGATAGCTGAA 2049
 DB 481 GTTATTGGCTATTTCTGGTATGATTAACAATAATTTCTCTGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGAACATCCCAAGACATTCGGAAAGTGCATTTTACCCACAGGAGAAATCCGT 2109
 DB 541 GGAGAACATCCCAAGACATTCGGAAAGTGCATTTTACCCACAGGAGAAATCCGT 600
 QY 2110 GTAGCAAAAGCAGGATCCCTACCTTGTGAAATGGCTTATGTATATAAATGTCTACTAC 2169
 DB 601 GTAGCAAAAGCAGGATCCCTACCTTGTGAAATGGCTTATGTATATAAATGTCTACTAC 660
 QY 2170 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 2229
 DB 661 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCAGCGACAGGTAAT 720
 QY 2230 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGAGCCTTTTACACAGAA 2289
 DB 721 GCTGAGATTGGAATTAAGGACATTAATTCACACATTTGGAAGAGCCTTTTACACAGAA 780
 QY 2290 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATTAACAGGAGACATTAATACAC 2349
 DB 781 CACTGGCTGTTCGATATATAAGTAAAGACACCTGATTAACAGGAGACATTAATACAC 840
 QY 2350 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 2409
 DB 841 AAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAAGAGACTACCAA 900
 QY 2410 AGGAGCTGTCTACATTAATAAGTAAAGTGGTTTAAAGAGGCAAGAAATATCTAG 2469
 DB 901 AGGAGCTGTCTACATTAATAAGTAAAGTGGTTTAAAGAGGCAAGAAATATCTAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972

RESULT 11

AF132552
 LOCUS AF132552 2417 bp mRNA linear INV 12-FEB-2003
 DEFINITION Drosophila melanogaster GM01839 full insert cDNA.
 ACCESSION AF132552

Db 301 GGTCCTTTCTCTTGTGATCTACATATTTGTATGTACCTCCAGAGGCTATGTT 360
Qy 1293 CTGCATCAAAATATCAACGATGAAGAGATTTTGTCTATATGCAATCAGTCTCT 1352
Db 361 CTGCATCAAAATATCAACGATGAAGAGATTTTGTCTATATGCAATCAGTCTCT 420
Qy 1353 CTACTTTCTGAGTGTGTCGACGATGTTGACTTTGACTTTCAGTCCAGTGTGTATGCT 1412
Db 421 CTACTTTCTGAGTGTGTCGACGATGTTGACTTTGACTTTCAGTCCAGTGTGTATGCT 480
Qy 1413 GTCTGCAATTCCTTTTCAATGTTTTTGGGACTATTTGGGGATGACATGAAAAGGGA 1472
Db 481 GTCTGCAATTCCTTTTCAATGTTTTTGGGACTATTTGGGGATGACATGAAAAGGGA 540
Qy 1473 AAATCCACCTGTGAGGACAGCTGTGATGAGGATGACAAAGAAACCAAGGAATTTGTA 1532
Db 541 AAATCCACCTGTGAGGACAGCTGTGATGAGGATGACAAAGAAACCAAGGAATTTGTA 600
Qy 1533 TGATAAGGACGTTAAAGTGAGGAACAATGCAACTGACAGGAAACCACTGAAGAGGGATT 1592
Db 601 TGATAAGGACGTTAAAGTGAGGAACAATGCAACTGACAGGAAACCACTGAAGAGGGATT 660
Qy 1593 AGGCTTAATATAAAGGATGTCACCATGTTGATGCTGATGCTATTTGATGATGCTTTC 1652
Db 661 AGGCTTAATATAAAGGATGTCACCATGTTGATGCTGATGCTATTTGATGATGCTTTC 720
Qy 1653 TGTCCACTGTACCTGGTCAAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCGCTC 1712
Db 721 TGTCCACTGTACCTGGTCAAGCAATGCTACTCTAGTCCAAAGTGTAGTCTGCGCTC 780
Qy 1713 ATCAATCATGATGCAACGAGGATATCTTAGATGATTTAGAGAGCTTACTTTGGCT 1772
Db 781 ATCAATCATGATGCAACGAGGATATCTTAGATGATTTAGAGAGCTTACTTTGGCT 840
Qy 1773 AAGCAAAATACAGATGAACATGCAAGTAATGCTTGTGTGGATTTATGGCTATCAGAT 1832
Db 841 AAGCAAAATACAGATGAACATGCAAGTAATGCTTGTGTGGATTTATGGCTATCAGAT 900
Qy 1833 AGCTGGATGCTTAATAGACTAGTGTGGGATATACACCTGGAATAACAGCCACAT 1892
Db 901 AGCTGGATGCTTAATAGACTAGTGTGGGATATACACCTGGAATAACAGCCACAT 960
Qy 1893 AGCACTGTGGGAAAGCTATGCTTCTAATGAAGACAGAGCTATATAAATCATGAGAC 1952
Db 961 AGCACTGTGGGAAAGCTATGCTTCTAATGAAGACAGAGCTATATAAATCATGAGAC 1020
Qy 1953 TCTAGATGATATGTTTGTGTTATTTTGGAGGGTATTTGGCTATTTCTGTGATGA 2012
Db 1021 TCTAGATGATATGTTTGTGTTATTTTGGAGGGTATTTGGCTATTTCTGTGATGA 1080
Qy 2013 TATCAACAAATTTCTTGGATGTTAGGATAGCTGAAGGAGAACATCCCAAGACATTCG 2072
Db 1081 TATCAACAAATTTCTTGGATGTTAGGATAGCTGAAGGAGAACATCCCAAGACATTCG 1140
Qy 2073 GGAAGTGAATTTTATCCCAAGGAGAAATTCGGTGTAGACAAAGAGGATCCCTTAC 2132
Db 1141 GGAAGTGAATTTTATCCCAAGGAGAAATTCGGTGTAGACAAAGAGGATCCCTTAC 1200
Qy 2133 TTTGTGAAATGCTTTATGATATAAATGTCATATCAGATTTGGAGAAATGCACTGGA 2192
Db 1201 TTTGTGAAATGCTTTATGATATAAATGTCATATCAGATTTGGAGAAATGCACTGGA 1260
Qy 2193 TTTTGTGACACCCGAGTTTGGACGACACATGATGCTGAGATTTGGAGAAATGCACTGGA 2252
Db 1261 TTTTGTGACACCCGAGTTTGGACGACACATGATGCTGAGATTTGGAGAAATGCACTGGA 1320
Qy 2253 TAAATTCAAATTTGGAAGAGCTTTTATCAGAACTGCTGCTTTGTTAGGATATATAA 2312
Db 1321 TAAATTCAAATTTGGAAGAGCTTTTATCAGAACTGCTGCTTTGTTAGGATATATAA 1380
Qy 2313 AGTAAAGCAGCTGATTAACAGGAGACATTTAGATCAAACTCCAGTCCAGCAATTTT 2372
Db 1381 AGTAAAGCAGCTGATTAACAGGAGACATTTAGATCAAACTCCAGTCCAGCAATTTT 1440

Qy 2373 CCCAAACAGAAAGTATTTGTCAAGAAGACTACCAAAAGGAGGCTGCTACATTAATAA 2432
Db 1441 CCCAAACAGAAAGTATTTGTCAAGAAGACTACCAAAAGGAGGCTGCTACATTAATAA 1500
Qy 2433 TAAGCTGCTTTTAAAGAAAGGCAAGAAATATCTTAAAGAGACTGTTTAA 2481
Db 1501 TAAGCTAGTGTTTTAAAGAAAGGCAAGAAACCTCTTAAAGAGACTGTTTAA 1549
RESULT 9
AX099510
LOCUS 2546 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 150 from Patent WO0119988.
ACCESSION AX099510
VERSION AX099510.1 GI:13538588
KEYWORDS
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Jacobs, K.; Mccoy, J.M.; Lavallie, E.R.; Collins-Racie, L.A.; Evans, C.;
Merberg, D.; Treacy, M.; Bowman, M.R.; Spaulding, V. and Agostino, M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 150 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES
source 1..2546
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 39.2%; Score 972; DB 6; Length 2546;
Best Local Similarity 100.0%; Pred. No. 5.4e-182;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1510 ABAAGAAACCAAGGAAATTTGTATGATAGGCAAGTAAAGTGAGCAACATGCAACTGAA 1569
Db 1 ABAAGAAACCAAGGAAATTTGTATGATAGGCAAGTAAAGTGAGCAACATGCAACTGAA 60
Qy 1570 CAGGAAAAAATCAAGAGGATTTAGGCTTAAATATAAAGCAATTCACCATGTTGATG 1629
Db 61 CAGGAAAAAATCAAGAGGATTTAGGCTTAAATATAAAGCAATTCACCATGTTGATG 120
Qy 1630 CTGATCTATTCATGATGTTTCTGTCACCTGTTACCTGGSTCACAAGCAATTCCTACTCT 1689
Db 121 CTGATCTATTCATGATGTTTCTGTCACCTGTTACCTGGSTCACAAGCAATTCCTACTCT 180
Qy 1690 AGTCCAAGTGTAGTCTGCTCATACAATCATGATGSCACCAGGAATATCTTAGATGAT 1749
Db 181 AGTCCAAGTGTAGTCTGCTCATACAATCATGATGSCACCAGGAATATCTTAGATGAT 240
Qy 1750 TTTTAGAGAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCACGATTAATGCT 1809
Db 241 TTTTAGAGAGCTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCACGATTAATGCT 300
Qy 1810 TGGTGGGATTAATGGCTATCAGATAGCTGGAATAGAACTACCTTGGTGGATAAT 1869
Db 301 TGGTGGGATTAATGGCTATCAGATAGCTGGAATAGAACTACCTTGGTGGATAAT 360
Qy 1870 AACACCTGGAATAACAGCCACATAGCTGTTGGGAAAGCTATGCTTCTTAATGAAACA 1929
Db 361 AACACCTGGAATAACAGCCACATAGCTGTTGGGAAAGCTATGCTTCTTAATGAAACA 420
Qy 1930 GCAGCCTATAAATCATGAGGACTCTAGATGATGATTTTGGTTATTTTGGAGGG 1989
Db 421 GCAGCCTATAAATCATGAGGACTCTAGATGATGATTTTGGTTATTTTGGAGGG 480
Qy 1990 GTTATTGGCTATCTGGGTGATGATTAACAATTTCTCTGATGCTTGGATAGCTGAA 2049
Db 481 GTTATTGGCTATCTGGGTGATGATTAACAATTTCTCTGATGCTTGGATAGCTGAA 540

```

1288 CTTGATGTCACCTCCAGCAGCCCTTGTTCTGATCAAAAATATCAACGATGAAGA 1347
|||||
Qy 1321 GTATTCTTGTCT 1332
|||||
Db 1348 GTATTCTTGTCT 1359
|||||

RESULT 8
LOCUS BC003206 1734 bp mRNA linear ROD 12-NOV-2003
DEFINITION Mus musculus RIKEN cDNA 1300006C19 gene, mRNA (cDNA clone
IMAGE:3587254), partial cds.
ACCESSION BC003206
VERSION BC003206.1 GI:13096823
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1734)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, R., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, J., Smal, D.E.,
Schnurch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2388257
12477932
2 (bases 1 to 1734)
Strausberg, R.
Direct Submission
Submitted (20-PEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC) Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseg, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LINL at: http://image.lnl.gov
Series: IRAX Plate: 10 Row: 0 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.
Location/Qualifiers

```

```

source 1. 1734
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="IMAGE:3587254"
/tissue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clone) outgrowth infected with the virus MMTV."
(clone_lib="NCI CGAP_Lu29"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
<1. 1734
/genes="1300006C19Rik"
/notes="synonym: Simp"
/db_xref="LocusID:68292"
/db_xref="MGI:1915542"
<1. 1549
/genes="1300006C19Rik"
/codon_start=2
/product="1300006C19Rik protein"
/protein_id="AAH03206.1"
/db_xref="GI:13096824"
/db_xref="LocusID:68292"
/db_xref="MGI:1915542"
/translation="GFOPIRSEHVAAGVFAALQAVAFLOYLRLDLTKOEFOTLFFL
GVSLAGAVFSLVLYLTGYIAPSGGFVSLMDTGYAKIHIFIASVSEHPTVVS
FFDHLILVCTFPAGLWFCIKNINDEKRVFVALVSAVYAGVAVRMLTLTIVVCM
SAIPNSFHYLGDMPKRENPVEDSDDEKRNPNGLYDKAGVKRKHVTEKEEE
GLPNKISIVTMLMLLMFPAVETWTNSVSSVSLASVYHNGDTRNLDDFREA
YFWRQNTDEHARVNMWDYGYQIAGNARTLLVNNWNSHIALVGRKMSNETAA
VKIMRSLDVLYLVIFGVGYSDDINDKFLMWVRIAEGRPKDIREGDFYQQGEFR
VDKAGPTLLNCLMYKSYVRFGEMLDFTPGFDRTRNAEIGNDKIFKHLEEAFT
SEHWLAVLYKYKAPDNRETLGKPRVNIIVPKQYLSKTTKRGVYVKNLVEKXGK
KTSKATV"
misc_feature 1. 1249
/genes="1300006C19Rik"
/notes="STT3; Region: Oligosaccharyl transferase STT3
subunit. This family consists of the oligosaccharyl
transferase STT3 subunit and related proteins. The STT3
subunit is part of the oligosaccharyl transferase (Oase)
complex of proteins and is required for its activity.
Oase transfers a lipid-linked core-oligosaccharide to
selected asparagine residues in the ER"
/db_xref="CDD:p1am02516"

ORIGIN
Query Match 52.7%; Score 1307.4; DB 10; Length 1734;
Best Local Similarity 90.3%; Pred. No. 1.9e-248;
Matches 1398; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 933 GGGATTCCAGCAATCAGAACAGTGAACACATGGCAGCTGCAGTGTCTTTCGATGCT 992
Db 1 GGGATTCCAGCAATCAGAACAGTGAACACATGGCAGCTGCAGTGTCTTTCGATGCT 60
Qy 993 GCAAGCTTATGTTCTTTCGAGTATCTGAGACAGCCAGTAAACAAAACAGAGTTCCAGAC 1052
Db 61 GCAAGCTTACGCTTTTTCGAGTATCTGAGACAGCGTTGACAAAACAGAGTTCCAGAC 120
Qy 1053 CTTTTTTTGGGTGTATCACTAGCTGCAGGTGCTGTGTCTTCTAGTGTCACTATTT 1112
Db 121 CTTTTTTTGGGTGTCTCACTAGCTGCAGGGCGGTGTGTCTTGTATGTATCTATCT 180
Qy 1113 GACTTATACAGTTTACATTTGACCATGGAGTGGCAGGTTTATTTTCATTGTGGGATCTGG 1172
Db 181 GACTATACAGTTTATATTTGACCATGGAGTGGCAGGTTTATTTTCATCTGGTACTGG 240
Qy 1173 GTATGAAAATACACATTCCTCAATTTTGCATCAGTGTCTGAGCATCACTAGCTGTTG 1232
Db 241 GTATGAAAATACACATTCCTCAATTTTGCATCAGTGTCTGAGCATCACTAGCTGTTG 300
Qy 1233 GGTGTCTTTCTTTTGTATCTACATATTTCTTGTATGTACCTTCCAGCAGCCCTTGGTT 1292

```


QY 2281 ACATCAGAACACTGGCTGTTAGTATATAAAGTAAAGACCTGATACAGGAGACA 2340
 Db |||||
 QY 2348 ACATCAGAACACTGGCTGTTAGTATATAAAGTAAAGACCTGATACAGGAGACA 2407
 Db |||||
 QY 2341 TTAGATCAACACCTCGAGTACCAACATTTTCCAAAACAGAGATTTTGTCAAGAAG 2400
 Db |||||
 QY 2408 CTAGGTCAACACCTCGAGTACCAACATCGTCCCAACAGAGATTTTGTCAAGAAG 2467
 Db |||||
 QY 2401 ACTACAAAAGGAGCGTGGCTACATTAATAAAGCTGCTTTTAAAGAAAGCAAGAAA 2460
 Db |||||
 QY 2468 ACTACTAAAGGAGCGTGGCTAGCTTTAAATAAAGCTAGTGTTTAAGAAAGCAAGAAG 2527
 Db |||||
 QY 2461 ATATCTAAGAGACTGTTTAA 2481
 Db |||||
 QY 2528 ACCTCTAAGAGACTGTTTAA 2548
 Db |||||
 RESULT 5
 CQ782554
 LOCUS 1664 bp DNA linear PAT 17-MAR-2004
 DEFINITION Sequence 2694 from Patent EP1396543.
 ACCESSION CQ782554
 VERSION CQ782554.1 GI:45502338
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Negai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 Primers for synthesizing full length cDNA clones and their use
 Patent: EP 1396543-A 2694 10-MAR-2004;
 Research Association for Biotechnology (JP)
 Location/Qualifiers
 1..1664
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 28..11362
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAP85831.1"
 /db_xref="GI:45502338"
 /translation="MARSPAPESKHXSLSNPSWGLMALGNSRHGHGPGCAQCAHKA
 AGKAPKAPAGLSQFAGWQSLSTLFLAWLAGFSRLFAVIRPESIHFE
 DPWNYRSTHLLSHGFYELNFWDERAWYPLGRIVGGTYPLMFTAGLIHILNTL
 NITVHIDVCVFLAPTFSGLSISTFLRELWNOGALLAACFIAIPGYISRSVAG
 SFDNEGIAIFALQPTYYLWVKSVKTSVFTWCCLSYFYMSWAMGYVFIINLPLH
 VFVLLMQRYSKRVYIAYSTFYIVGLILSMQIPFVGFPQIRTSHEHMAAGVPAALQAY
 AFLQVLRDLTKQBFOTLFFLVSLAAGAVFLSVLYLTVTGVIAPWSGRFVSLWDTGY
 AKIHPIIASVSEHQPPTWVSFFFDLHLVCTFPAGLMFCIKNINDERVGA"
 ORIGIN
 Query Match 53.6%; Score 1328.8; DB 6; Length 1664;
 Best Local Similarity 99.8%; Pred. No. 1.1e-252;
 Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGCGGAGCCCTGGCGCCCGGAGAGCAAGCAGAGTCGTCCTCAACTCGTCCCGCTGG 60
 Db |||||
 QY 28 ATGGCGGAGCCCTGGCGCCCGGAGAGCAAGCAGAGTCGTCCTCAACTCGTCCCGCTGG 87
 Db |||||
 QY 61 AGTGCCTCATGCGCTTGGGAAAACAGCGGCGCAGGCCACCAACCGGCGCGGCGCCAGTGC 120
 Db |||||
 QY 88 AGTGCCTCATGCGCTTGGGAAAACAGCGGCGCAGGCCACCAACCGGCGCGGCGCCAGTGC 147
 Db |||||
 QY 121 GGGCAACAGCGGCGGCG 180
 Db |||||
 QY 148 GGGCAACAGCGGCGGCG 207
 Db |||||
 QY 181 GGGCTGTGCGACCGCGCTGGGTGGCAGTGGCTTCTCTCTTCAACATCTCTCTCTCTCT 240
 Db |||||

Db 208 GGGCTGTGCGACGGCGTGGGTGGCAGTCGCTTCTCTCTTCAACATCTCTCTCTCTCTG 267
 QY |||||
 Db 241 TGGCTTGGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 300
 Db 268 TGGCTTGGCGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 327
 QY |||||
 Db 301 GAGTTCGACCCGCTGTTAACTATAGATCAACACATCATCTTTCATCTCTCATGGTTCAT 360
 Db 328 GAGTTCGACCCGCTGTTAACTATAGATCAACACATCATCTTTCATCTCTCATGGTTCAT 387
 QY |||||
 Db 361 GAATTTTAAATTTGTTTGGTGAAGAGCATGGTATCCACTAGGAGAAATAGTAGTGGT 420
 Db 388 GAATTTTAAATTTGTTTGGTGAAGAGCATGGTATCCACTAGGAGAAATAGTAGTGGT 447
 QY |||||
 Db 421 ACTGTTTACCAGGGTGTGATPAAACCGCTGGCTTATTCATTGGATTTTAAATACATG 480
 Db 448 ACTGTTTACCAGGGTGTGATPAAACCGCTGGCTTATTCATTGGATTTTAAATACATG 507
 QY |||||
 Db 481 AACATTAACCTTCCATTAAGAGACGTATGTGTCTCTTGGCACCACTTTTATAGCGGCTT 540
 Db 508 AACATTAACCTTCCATTAAGAGACGTATGTGTCTCTTGGCACCACTTTTATAGCGGCTT 567
 QY |||||
 Db 541 ACATCTATATCTACTTCTCTCTTCAAGAGAACTTTTGGAAACCAAGGAGCAGGACTTTTA 600
 Db 568 ACATCTATATCTACTTCTCTCTTCAAGAGAACTTTTGGAAACCAAGGAGCAGGACTTTTA 627
 QY |||||
 Db 601 GCTGCTTGTGTTTATTTGCTATTTGACAGGCTACATATCTCGGTCACTAGTGGATCTT 660
 Db 628 GCTGCTTGTGTTTATTTGCTATTTGACAGGCTACATATCTCGGTCACTAGTGGATCTT 687
 QY |||||
 Db 661 GATAATGAAGGCAATCTGCTATTTTGGCTTCACTAGTTCACATATTTATGGGTAAAAATCT 720
 Db 688 GATAATGAAGGCAATCTGCTATTTTGGCTTCACTAGTTCACATATTTATGGGTAAAAATCT 747
 QY |||||
 Db 721 GTAAAACTGGGTCAGTCTTTTGGCAATGCTGCTGCTATCTCTTCTATATGTC 780
 Db 748 GTAAAACTGGGTCAGTCTTTTGGCAATGCTGCTGCTATCTCTTCTATATGTC 807
 QY |||||
 Db 781 TCTGCTTGGGTCGTTATGTAATTTATCAATCTTATTTCCACTGATGATTTGTTG 840
 Db 808 TCTGCTTGGGTCGTTATGTAATTTATCAATCTTATTTCCACTGATGATTTGTTG 867
 QY |||||
 Db 841 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTTCTACATTTG 900
 Db 868 TTACTGATGAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTTCTACATTTG 927
 QY |||||
 Db 901 GGTTTAATATTTCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACCAAGTGA 960
 Db 928 GGTTTAATATTTCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACCAAGTGA 987
 QY |||||
 Db 961 CACATGCGACCTGCGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020
 Db 988 CACATGCGACCTGCGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1047
 QY |||||
 Db 1021 AGAGACCGATTAACAAAACAGAGTTCAGACCTTTCTTTTGGGTGATFACATAGCT 1080
 Db 1048 AGAGACCGATTAACAAAACAGAGTTCAGACCTTTCTTTTGGGTGATFACATAGCT 1107
 QY |||||
 Db 1081 GCAGGTGCTGCTGTTCTTGTAGTGTCTATTTGACTTATACAGGTTTACATTGCACATG 1140
 Db 1108 GCAGGTGCTGCTGTTCTTGTAGTGTCTATTTGACTTATACAGGTTTACATTGCACATG 1167
 QY |||||
 Db 1141 AGTGGCAGGTTTATTCATTTGGGATCTGGGTATGCAAAATACATTCATCAATTTAT 1200
 Db 1168 AGTGGCAGGTTTATTCATTTGGGATCTGGGTATGCAAAATACATTCATCAATTTAT 1227
 QY |||||
 Db 1201 GCATCAGTGTCTGAGCATCAACCTAGCTTGGGTGCTTTCTTTGATCTACATATT 1260
 Db 1228 GCATCAGTGTCTGAGCATCAACCTAGCTTGGGTGCTTTCTTTGATCTACATATT 1287
 QY |||||
 Db 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGTGATCAAAATATCAACATCAAGTGAAGA 1320
 Db 1288 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGTGATCAAAATATCAACATCAAGTGAAGA 1347

Db 137 AGCGGCTCATGGCTCTGGGAAACAGCGCCACAGCGGCACCATGGGCCCGGAAACCCAGAGC 196
Qy 121 CGGCACAAGCGCGGG 180
Db 197 GCGTCCAGGGGG-----GCGGCGCCGAAAGCCGGGGGGGGGGGGGGGGGGGGGGGGGG 247
Qy 181 GGGGTGTGCGACGCGGGGTGGGTGCGAGTCGCTCTCTCTTCAACATCTCTTCTCTCTGGCC 240
Db 248 GGCCTGTGCGACGCGGG 307
Qy 241 TGGCTTGGCGGCTTCACTGCGGCTTTCGCGGTCATTCGCGCTTCGAAAGCATTCAC 300
Db 308 TGGCTTGGCGGCTTCACTGCGGCTTTCGCGGTCATTCGCGCTTCGAAAGCATTCAC 367
Qy 301 GAGTTCGACCGGCTTAACTATAGATCAACACATCATCTTGCATCTCATGGATCTAT 360
Db 368 GAGTTCGACCGGCTTAACTATAGATCAACACATCATCTTGCATCTCATGGATCTAT 427
Qy 361 GAATTTTAAATGGTTGATGAAGAGCATGTATCCACTAGGAAGATAGTGGTGT 420
Db 428 GAGTTTCTAAATGGTTGATGAAGAGCATGTATCCACTAGGAAGATAGTGGTGGC 487
Qy 421 ACTGTTTACCGAGGTTGATGATAACCGCTGGGCTTATTCATTTGGAATTTAAATACATG 480
Db 488 ACCGTTTACCGAGGTTGATGATAACAGCTGGGCTTATTCATTTGGAATTTAAATACATG 547
Qy 481 AACATACGTTTCAATGAAGAGCATGTATGGTTCCTTGCACCACTTTTACGGGCTT 540
Db 548 AACATAACAGTTTCAATGAAGAGCATGTATGGTTCCTTGCACCACTTTTACGGGCTT 607
Qy 541 ACATCTATCTACTTTCCTTCTTCAAGAGAACTTTGGAACCAAGAGGAGGAGCTTTTA 600
Db 608 ACATCTATCTACTTTCCTTCTTCAAGAGAACTTTGGAACCAAGAGGAGGAGCTTTTA 667
Qy 601 GCTGCTGTTTATTTGCTATTTGACAGGCTACATATCTGGTTCAGTACGTAGTGCCTTT 660
Db 668 GCTGCTGCTTCTATTTGCTATTTGACAGGCTACATATCTGGTTCAGTACGTAGTGCCTTT 727
Qy 661 GATATGAAGGCTTCTATTTTTCAGTTCAGTTCACATATTTATTTGGTAAATCT 720
Db 728 GATATGAAGGCTTCTATTTTTCAGTTCAGTTCACATATTTATTTGGTAAATCT 787
Qy 721 GTAAATCTGGTCACTTTTGGACAAATGTGCTGCTTATCTTATTTCTATATGTC 780
Db 788 GTGAAGACCGGCTGTGTTCTGGACAAATGTGCTGCTTATTTCTATATGTC 847
Qy 781 TCTGCTGGGGTGTATGTATTTATCATCAATCTTATTCCTGCTGCTATTTTGTGTTG 840
Db 848 TCTGCTGGGGGTTATGTATTTATCATCAATCTTATTCCTGCTGCTATTTTGTGTTG 907
Qy 841 TTACTGTGAGATACAGAAAGAGCTTACATAGCATATAGCATTTCTACATGTC 900
Db 908 CTGCTGTGAGAGGTCAGCAAGAGAGTCTTACATAGCATATAGCATTTTGTACATGTC 967
Qy 901 GGTTTAATTTATCAATGAGATACCTTTTGTGGGATTCAGGCAATCAGAAACAGTCAA 960
Db 968 GGTTTAATTTATCAATGAGATACCTTTTGTGGGATTCAGGCAATCAGAAACAGTCAA 1027
Qy 961 CACATGGCAGTTCAGGTGCTTTGCAATTCGCAAGCTTATCTTTCTTGGAGTATCTG 1020
Db 1028 CACATGGCAGTTCAGGTGCTTTGCAATTCGCAAGCTTATCTTTCTTGGAGTATCTG 1087
Qy 1021 AGAGACCGATTTAAACAAACAGAGTTCAGACCGCTTTTCTTTTGGGTGTATCACTAGCT 1080
Db 1088 AGAGACCGGTTGACAAACAGAGTTCAGACCGCTTTTCTTTTGGGTGTATCACTAGCT 1147
Qy 1081 CGAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
Db 1148 GCAGGCGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1207
Qy 1141 AGTGGCAGGTTTATTTCTATTTGGGATCTGGGTATGCAAAATACACATTTCAATTTT 1200
Db 1208 AGTGGCAGGTTTATTTCTATTTGGGATCTGGGTATGCAAAATACACATTTCAATTTT 1267

Qy 1201 GCATCAGTCTCTGAGCATCAACTAGACTTGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1260
Db 1268 GCATCAGTCTCTGAAACATCAGCTTACGATGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1327
Qy 1261 CTTGTATGTACCTTCCAGCAGCGCTTTGGTTCGTGCTCAAAAAATATCAACGATGAAGA 1320
Db 1328 CTTGTATGTACCTTCCAGCAGCGCTTATGGTTCGTGCTCAAAAAATATCAACGATGAAGA 1387
Qy 1321 GTATTTGTGCTCTATATGCAATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
Db 1388 GTATTTGTGCTCTATATGCAATCAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1447
Qy 1381 ATGTTGACTTTGACTTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 1448 ATGTTGACTTCCAGCAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507
Qy 1441 GAGCATTTTGGGGATGACATGAAAGGAAATCCACTGTGGAGGACAGCAGTGTAT 1500
Db 1508 GAGCATTTTGGGGATGACATGAAAGGAAATCCACTGTGGAGGACAGCAGTGTAT 1567
Qy 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAGGAGTAAAGTGAGGAAACAT 1560
Db 1568 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAGGAGTAAAGTGAGGAAACAT 1627
Qy 1561 GCAACTGAAACAGGAAACCAAGGAGGATTTAGGCCCTTATATAAAAGCATTTGTCAAC 1620
Db 1628 GTGACAGAGAGAGGAAACCTTGAAGAGGCTTTGGGCCCAACATCAAAAGCATTTGTGACC 1687
Qy 1621 ATGTTGACTCTGATGCTATTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 1688 ATGTTGACTCTGATGCTATTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1747
Qy 1681 GCTTACTCTAGTCCAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1748 GCTTACTCTAGTCCAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1807
Qy 1741 TTAGATGATTTTAGAGAGCTTTACTTTTGGCTTAAAGGCAAAATACAGATGAACATGCAAGA 1800
Db 1808 TTAGATGATTTTAGAGAGCTTTACTTTTGGCTTAAAGGCAAAATACAGATGAACATGCAAGA 1867
Qy 1801 GTATGCTCTTGGTGGATTTATGCTATCAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db 1868 GTATGCTCTTGGTGGATTTATGCTATCAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1927
Qy 1861 GTGATTAATAACACTTGGAAATAACAGCAATAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1928 GTGATTAATAACACTTGGAAATAACAGCAATAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1987
Qy 1921 AATGAAACAGCAGCTTATAAATCATGAGGCTCTAGATGATGATGATGATGATGATGATGATGATGAT 1980
Db 1988 AATGAAACAGCAGCTTATAAATCATGAGGCTCTAGATGATGATGATGATGATGATGATGATGATGAT 2047
Qy 1981 TTTGGAGGCTTATTTGGCTTATTTCTGGTATGATATCAACAAATTTCTTCTGATGCTTAGG 2040
Db 2048 TTTGGAGGCTTATTTGGCTTATTTCTGGTATGATATCAACAAATTTCTTCTGATGCTTAGG 2107
Qy 2041 ATAGCTGAAGGAGAAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCCAAGGGA 2100
Db 2108 ATAGCTGAAGGAGAAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCCAAGGGA 2167
Qy 2101 GAATTCGGTGTAGACAAAGCAGGATCCCTTACTTTGTTGAAATTTGCTTATGTAATAAATG 2160
Db 2168 GAGTTCGGTGTAGACAAAGCAGGATCCCTTACTTTGTTGAAATTTGCTTATGTAATAAATG 2227
Qy 2161 TCATCTACAGATTTGGAGAAATGAGTGGATTTTCTGACACCCCAAGGTTTGAACGGA 2220
Db 2228 TCATCTACAGATTTGGAGAAATGAGTGGATTTTCTGACACCCCAAGGTTTGAACGGA 2287
Qy 2221 ACAGGTAATGCTGAGATTTGAGAAATAGGACATTTAAATTTCAACATTTGGAAGAGGCTTTT 2280
Db 2288 ACAGGTAATGCTGAGATTTGAGAAATAGGACATTTAAATTTCAACATTTGGAAGAGGCTTTT 2347

1197 CTGCTGATGACAGAGTACAGACAGAGTCTACATAGCATATACACCTTTTACATGTTG 1256
901 GGTTAATATTATCAATGACAGATACCTTTTGTGGGATCCAGCAATCAGAACAGTGA 960
1257 GGTTAATATTATCAATGACAGATACCTTTTGTGGGATCCAGCAATCAGAACAGTGA 1316
961 CACATGGCAGTGCAGGTGCTTTTGGCAATGCTGCAAGCTTATGCTTTTGTGAGTATCG 1020
1317 CACATGGCAGTGCAGGTGCTTTTGGCCTGCTGCAAGCTTATGCTTTTGTGAGTATCG 1376
1021 AGAGACCATTAACAAACAGAGTCTCCAGACCTTTTCTTTTGGGTGATCAGTACGT 1080
1377 AGAGACCGTTGACAAACAGAGTCTCCAGACCTTTTCTTTTGGGTGATCAGTACGT 1436
1081 GCAGGTGCTGCTTCTTGTAGTCACTTATTTGATCTATACAGTTTACATGACCAATGG 1140
1437 GCAGGCGCTGCTTCTTGTAGTCACTTATTTGATCTATACAGTTTATTTGACCAATGG 1496
1141 AGTGGCAGTTTATTCATTTGGGATCTGCGGTATGCAAAATACACATTCCAATATT 1200
1497 AGTGGCAGTTTATTCATTTGGGATCTGCGGTATGCAAAATACACATTCCAATATT 1556
1201 GCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTTTCTTTTGTGATCTACATATT 1260
1557 GCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTTTCTTTTGTGATCTACATATT 1616
1261 CTGTTGATGATCTTCCAGCAGGCTTTGCTGCTGATCAAAATATCAACGATGAAGA 1320
1617 CTGTTGATGATCTTCCAGCAGGCTTTGCTGCTGATCAAAATATCAACGATGAAGA 1676
1321 GTATTTGTTGCTTATATGCAATCAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
1677 GTATTTGTTGCTTATATGCAATCAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1736
1381 ATGTTGATCTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
1737 ATGTTGATCTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
1441 GAGCAGTATTTGGGGATGACATGAAGAGGAAATCCACCTGTGAGGACGAGTGTAT 1500
1797 GAGCAGTATTTGGGGATGACATGAAGAGGAAATCCACCTGTGAGGACGAGTGTAT 1856
1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAGGAAATCAAGTGAAGAAAT 1560
1857 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAAGGAGGAAATCAAGTGAAGAAAT 1916
1561 GCAACTGACAGGAAAGAAATGTAAGAGGAAATTTAGGCTTAAATATAAGGAAATTTGTCAC 1620
1917 GTGACAGGAAAGAAACCAAGGAAATTTAGGCTTAAATATAAGGAAATTTGTCAC 1976
1621 ATGTTGATCTGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
1977 ATGTTGATCTGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2036
1681 GCCTACTCTAGTCCAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
2037 GCCTACTCCAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2096
1741 TTAGATGATTTTACAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATCAACATGACAGA 1800
2097 TTAGATGATTTTACAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATCAACATGACAGA 2156
1801 GTATGCTTGGTGGGATTTATGCTATCAGATAGCTGGAATGCTTAAATGAATCAGTTG 1860
2157 GTCTATGCTGCTGGGATCTAGGCTATCAGATAGCTGGAATGCTTAAATGAATCAGTTG 2216
1861 GTGATATTAACACTCGAATAACAGGCAATAGCAGTGTGGAAGAAAGCTATGCTTCT 1920
2217 GTGATATTAACACTCGAATAACAGGCAATAGCAGTGTGGAAGAAAGCTATGCTTCT 2276
1921 AATGAACAGCAGGCTTATAAATCATAGGACTCTAGATGTAGATTTATTTTGTGTTATT 1980

2277 AATGAACCGCCCTTATAAATCATAGGTCCTTGATGTCGATTATGTGTGTTATT 2336
1981 TTTGGAGGGTATTATGCTTATTCGTGATGATATATCAAAATTTCTCTGATGTTAGG 2040
2337 TTTGGAGGAGTATGCTTATTCGGGGACGATATCAAAAGTCTCTCTGATGTTAGG 2396
2041 ATAGCTGAGGAGAAATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2100
2397 ATAGCTGAGGAGGAGATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2456
2101 GAATTCGCTGATAGCAAAAGCAGGATCCCTATCTTTGTTGAAATGCTTATGTATATAAATG 2160
2457 GAGTTCGAGTACAAAGCTGGTCTCTACTCTCTGTTAAATGCTTATGTATATAAATG 2516
2161 TCATATCAGATTTGAGGAAATGAGCTGAGTATTTGTTGAAATGCTTATGTATATAAATG 2220
2517 TCATATCAGATTTGAGGAAATGAGCTGAGTATTTGTTGAAATGCTTATGTATATAAATG 2576
2221 ACAGCTAATGCTGAGATTTGAAATAGGACATTAATTAATTCAAACATTTGGAAGAGCTTT 2280
2577 ACAGCTAATGCTGAGATTTGAAATAGGACATTAATTAATTCAAACATTTGGAAGAGCTTT 2636
2281 ACATCAGACACTGGCTTGTAGGATATATAAAGTAAAGACCTGATTAACAGGAGACA 2340
2637 ACATCAGACACTGGCTTGTAGGATATATAAAGTAAAGACCTGATTAACAGGAGACA 2696
2341 TTAGATCACAACCTCGAGTCAACCAATTTTCCCAAAACAGAAATTTTGTCAAAAGAG 2400
2697 CTAGGTCACAACCTCGAGTCAACCAATTTTCCCAAAACAGAAATTTTGTCAAAAGAG 2756
2401 ACTACCAAAAGAGGCTGCTGCTACATTAATAAATAGCTGTTTAAAGAAAGGCAAGAAA 2460
2757 ACTACTAAAGAGGCTGCTGCTACATTAATAAATAGCTGTTTAAAGAAAGGCAAGAAA 2816
2461 ATATCTAAGAGAGCTGTTTAA 2481
2817 ACCTCTAAGAGAGCTGTTTAA 2837

RESULT 4
AX799084 LOCUS 2710 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 3 from Patent WO03054008.
ACCESSION AX799084
VERSION AX799084.1 GI:37605059
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Perreault, C. and McBride, K.
TITLE Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
JOURNAL Patent: WO 03054008-A 3 03-JUL-2003;
COMPATIGENE Inc. (CA)
FEATURES
Location/Qualifiers
source 1. .2710
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

ORIGIN
Query Match 83.8%; Score 2078.6; DB 6; Length 2710;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 234; Indels 9; Gaps 1;
QY 1 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTCTGCTCCTCAACTCGTCCCGG 60
DB 77 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTCTGCTCCTCAACTCGTCCCGG 136
QY 61 AGTGGCTCATGCTCCTTGGGAAACAGCGGCAAGCCAGCCCGGCGGCCCGAGTGC 120

Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences (26), 16899-16903 (2002)
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 MEDLINE
 12477932
 PUBLISHED
 2 (bases 1 to 4236)
 STRAUSBERG, R.
 DIRECT SUBMISSION
 Submitted (15-MAY-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC) Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
 Thomas L. Casavant.
 Web site: <http://genome.uiowa.edu>
 Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
 Bonaldi, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
 Fishler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
 Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
 Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: Plate: Row: Column: 0
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis. Similarity but not identity to protein.

FEATURES

source

1..4236
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="MGC:64679 IMAGE:6837097"
 /tissue="Brain, enriched mouse brain 12.5dpc"
 /clone_lib="NIH BMAP_F10"
 /lab_host="DH10B"
 /note="Vector: pYX-ASC"
 1..4236
 /gene="1300006C19Rik"
 /note="synonym: Simp"
 /db_xref="LocusID:68292"
 /db_xref="WGI:1915542"
 366..2837
 /gene="1300006C19Rik"
 /codon_start=1
 /product="1300006C19Rik protein"
 /protein_id="AAH52433.1"
 /db_xref="GI:30851502"
 /db_xref="LocusID:68292"
 /db_xref="WGI:1915542"
 /translation="MAEPSPAPESKXSLNSPQGLMNLGNRHHGHGRTQASRA
 AAFPGPAGLSGLSQPAGNQLSFTILFLANLAGFSRLFAVIREGTHIEFDPE
 FNKSTHLSHSGFYFLNFDERAWYPLGRIVGGTVYPLGLMTHLHMLINLIT
 VHRDVCVLAFTFSLTSISTELLRELNQAGLLAACFIAIIVGYISRVAGSFD
 NEGIAIPALQFTLYLVKYSKVTGFWTCCCLSYFVWAWGGVYFVILNIPHFV
 LLLMQRYSKRYTAYSTFYVILGSLMQIPFVGFQPIRTSEHMAAGVFAILLQYAF
 QYLRDLTKQFOTFLPLGSLAAGAVLSVLTVTGVIAFWSGRFYSLMDTGAKI
 HIPIASVSHQRTWVNSRPFDLHILVCTFPAGLMFCIKINDERFVALVAISVY
 AGVMYLMILTLTPVCMNSAIAESNVEHLDGDMKRENPVEDSDSDKRNPNLY
 DKAGYKRVTEKEKEPEGGPNKISIVTLMMLLMFMFAVCTTSDNSYSSSVL
 ASYNDGTRNILDDEAFYFWLQNTDEHARVNSWMDYVYIAGVAGVITLVDNWN
 NSHIALYKXKASNEAAKYIMRSLDVLVIFGGVIGYSGDDINKFLMWVIAEGE
 HPKDIREGDTYQGGFPRVDKAGSPILLNCLMYKSYRFGEMQJDFTPFPDRTN
 AEIGNKIDIKFHLSEAPTSEHNLVRIYKVKAPDNRETLGHKPRVTNIVPKQYLSKT
 TKRGRGVVKNKLVFKGKTKTSKTV"

gene

CDS

misc_feature 567..2537
 /gene="1300006C19Rik"
 /note="STT3. Region: Oligosaccharyl transferase STT3
 subunit. This family consists of the oligosaccharyl
 transferase STT3 subunit and related proteins. The STT3
 subunit is part of the oligosaccharyl transferase (Oase)
 complex of proteins and is required for its activity.
 Oase transfers a lipid-linked core-oligosaccharide to
 selected asparagine residues in the ER"
 /db_xref="CDD:pfam02516"

ORIGIN

Query Match 83.9%; Score 2081.8; DB 10; Length 4236;
 Best Local Similarity 90.3%; Pred. No. 0;
 Matches 2240; Conservative 0; Mismatches 232; Indels 9; Gaps 1;
 QY 1 ATGGCGAGCCCTCGGCGCCCGGAGAGCAAGTCCTCCCTCAACTCGTCCCGCTGG 60
 DB ATGGCGAGCCCTCGGCGCCCGGAGAGCAAGTCCTCCCTCAACTCGTCCCGCTGG 425
 QY 61 AGTGGCTCATGCGCTTGGGAAACAGCGGACAGGCGCACACGCGGCGCGGCGCGCTGC 120
 DB AGCGGCTCATGCGCTTGGGAAACAGCGGACAGGCGCACACGCGGCGCGGCGCGCTGC 485
 QY 121 GGCACAAAGCGCGGCG 180
 DB GCGTCCAGGGCG-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 536
 QY 181 GGCTGTGCGAGCGCGCTGGGCGAGTCGCTTCTCTTCACATCTCTTCCTCGGCC 240
 DB GGCTGTGCGAGCGCGCTGGGCGAGTCGCTTCTCTTCACATCTCTTCCTCGGCC 596
 QY 241 TGGCTTGGCGCTTCAGCTCGCGCTCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCG 300
 DB TGGCTTGGCGCTTCAGCTCGCGCTCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCG 656
 QY 301 GAGTTCGACCGCGCTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 360
 DB GAGTTCGACCGCGCTTAACTATAGATCAACATCATCTTGCATCTCATGGTTCAT 716
 QY 361 GAATTTTAAATGGTTTGTATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGT 420
 DB GAGTTTCTAAATGGTTTGTATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGT 776
 QY 421 ACTGTTTACCGAGGTTGATGATAACCGCTGGCTTATCATTCGATTTTAAATACATTG 480
 DB ACCGTTTACCGAGGTTGATGATAACCGCTGGCTTATCATTCGATTTTAAATACATTG 836
 QY 481 AACATAACTGTTTACATAAGAGACGATATGTGTGTTTCTTGCACCACTTTTAGCGGCTT 540
 DB AACATAAGAGTTTCAATAGAGATGTGTGTATTCCTTGCACCACTTTTAGCGGCTT 896
 QY 541 ACATCTATATCTACTTTCTCTTCAAGAGACCTTTTGGAAACCAAGAGAGAGGACTTTTA 600
 DB ACATCTATATCTACTTTCTCTTCAAGAGACCTTTTGGAAACCAAGAGAGAGGACTTTTA 956
 QY 601 GCTGCTTGTGTTTATGCTATTGTACAGGCTACATATCTCGGTGAGTAGTGGATCTCTT 660
 DB GCTGCTTGTGTTTATGCTATTGTACAGGCTACATATCTCGGTGAGTAGTGGATCTCTT 1016
 QY 661 GATAATGAAGGACTGCTATTTTTCGACTTCAGTTTCACTACTATTTATGGGTAATACT 720
 DB GATAATGAAGGACTGCTATTTTTCGACTTCAGTTTCACTACTATTTATGGGTAATACT 1076
 QY 721 GTAAAAACCTGGCTCAGTTTTTTTGGCAATGTGCTGCTTATCTATTTCTATATGGTC 780
 DB GTGAAGACCGGCTGCTGTTCTGGCAATGTGCTGCTTATCTATTTCTATATGGTC 1136
 QY 781 TCTGCTTGGGCTGCTTATGATTTATTCATCAATCTTATTCACATGATTTTGTGTTG 840
 DB TCTGCTTGGGCTGCTTATGATTTATTCATCAATCTTATTCACATGATTTTGTGTTG 1196
 QY 841 TTACTGATGACAGATACAGCAAAAGAGTCTACATAGCATATAGCATTTCTACATTTGTG 900

Db 781 TCTGTTGGGTTGTTATGTTATTTATCATCAATCTTATTCACATGATGTTGTTG 840
Qy 841 TTACTGATGCGAGATACAGCAAAAGAGTCTACATGACATATAGCATCTTCTACATGTTG 900
Db 841 TTACTGATGCGAGATACAGCAAAAGAGTCTACATGACATATAGCATCTTCTACATGTTG 900
Qy 901 GGTTTAAATATTAATCAATGACATACCTTTTGTGGGATCCAGCAATCAGAAACAAGTAA 960
Db 901 GGTTTAAATATTAATCAATGACATACCTTTTGTGGGATCCAGCAATCAGAAACAAGTAA 960
Qy 961 CACATGCGAGCTGCGAGTGTCTTTCATCTGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020
Db 961 CACATGCGAGCTGCGAGTGTCTTTCATCTGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020
Qy 1021 AGAGACCGATTAACAAAACAAGAGTTCAGACCCCTTTTCTTTTGGGTGTATCACTAGCT 1080
Db 1021 AGAGACCGATTAACAAAACAAGAGTTCAGACCCCTTTTCTTTTGGGTGTATCACTAGCT 1080
Qy 1081 GCAGGTGCTGTTCCCTTTAGTGTCTATCTTATTTGACTTATACAGTTACATGTCACCATGG 1140
Db 1081 GCAGGTGCTGTTCCCTTTAGTGTCTATCTTATTTGACTTATACAGTTACATGTCACCATGG 1140
Qy 1141 AGTGCAGGTTTATTTATCTATTGTTGGGATCTGGGTATGCAAAAATACACATTCCTTAA 1200
Db 1141 AGTGCAGGTTTATTTATCTATTGTTGGGATCTGGGTATGCAAAAATACACATTCCTTAA 1200
Qy 1201 GCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTTGATCTACATATTT 1260
Db 1201 GCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTTGATCTACATATTT 1260
Qy 1261 CTTGTATGACCTTCCAGCAGGCTTTGTTCTGCAATCAAAAATACACATTCCTTAA 1320
Db 1261 CTTGTATGACCTTCCAGCAGGCTTTGTTCTGCAATCAAAAATACACATTCCTTAA 1320
Qy 1321 GTATTTGTTGCTATATGCAATCAGTCTGCTGCTTCTTCTGAGTGATGTTGCACTG 1380
Db 1321 GTATTTGTTGCTATATGCAATCAGTCTGCTGCTTCTTCTGAGTGATGTTGCACTG 1380
Qy 1381 ATGTTGACTTTGACTCCAGTCTGTTGTTATGCTGCTGCTTCTTCTTCAATGTTT 1440
Db 1381 ATGTTGACTTTGACTCCAGTCTGTTGTTATGCTGCTGCTTCTTCTTCAATGTTT 1440
Qy 1441 GAGCCTATTTGGGGATGACATGAAAGGGAATCCACTGTTGGAGGACAGCTGAT 1500
Db 1441 GAGCCTATTTGGGGATGACATGAAAGGGAATCCACTGTTGGAGGACAGCTGAT 1500
Qy 1501 GAGGATGCAAAAGAAACCAAGGAAATTTGTATGATAGGAGGATTAAGTGAGGAAACAT 1560
Db 1501 GAGGATGCAAAAGAAACCAAGGAAATTTGTATGATAGGAGGATTAAGTGAGGAAACAT 1560
Qy 1561 GCACTGAAACAGGAAACCAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGATTA 1620
Db 1561 GCACTGAAACAGGAAACCAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGATTA 1620
Qy 1621 ATGTTGATGCTGATGCTATTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
Db 1621 ATGTTGATGCTGATGCTATTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
Qy 1681 GCTTACTCTAGTCCAAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Db 1681 GCTTACTCTAGTCCAAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Qy 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTTAAAGCAATGACATGACATGACATGACATG 1800
Db 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTTAAAGCAATGACATGACATGACATGACATG 1800
Qy 1801 GTAAATGCTTTGGTGGATTTATGCTTATCAGATAGTGGGATGCTTAAAGCAATGACATGCTG 1860
Db 1801 GTAAATGCTTTGGTGGATTTATGCTTATCAGATAGTGGGATGCTTAAAGCAATGACATGCTG 1860
Qy 1861 GTGATTAATTAACCTGGATTAACAGGACATAGCAGCTGTTGGGAAAGCTATGCTTCTTCT 1920

Db 1861 GTGATTAATTAACCTGGATTAACAGGACATAGCAGCTGTTGGGAAAGCTATGCTTCTTCT 1920
Qy 1921 AATGAACAGCAGCCTTAAATATCATGAGGACTCTAGATGTAGATATGTTTGGTTATT 1980
Db 1921 AATGAACAGCAGCCTTAAATATCATGAGGACTCTAGATGTAGATATGTTTGGTTATT 1980
Qy 1981 TTTGGAGGGTTATTGCTTATTTCTGATGATATCAACAAATTTCTGATGATGTTAGG 2040
Db 1981 TTTGGAGGGTTATTGCTTATTTCTGATGATATCAACAAATTTCTGATGATGTTAGG 2040
Qy 2041 ATAGCTGAGGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGAA 2100
Db 2041 ATAGCTGAGGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGAA 2100
Qy 2101 GAATTCGCTGTAGACAAAGCAGGATCCCTTACTTTTGTGTAATTTGCTTATGATATAAATG 2160
Db 2101 GAATTCGCTGTAGACAAAGCAGGATCCCTTACTTTTGTGTAATTTGCTTATGATATAAATG 2160
Qy 2161 TCATATCTACAGATTTTGGAGAAATGCGACTGGATTTTCTACACCCCGAGGTTTGAACCGA 2220
Db 2161 TCATATCTACAGATTTTGGAGAAATGCGACTGGATTTTCTACACCCCGAGGTTTGAACCGA 2220
Qy 2221 ACAGCTAAATGCTGAGATTCGAAATTAAGGACATTAATTTCAACATTTTGAAGAGCCTTT 2280
Db 2221 ACAGCTAAATGCTGAGATTCGAAATTAAGGACATTAATTTCAACATTTTGAAGAGCCTTT 2280
Qy 2281 ACATCAGAACACCTGGCTTGTAGGATATATAAAGTAAAGACACCTGTATAACAGGAGACA 2340
Db 2281 ACATCAGAACACCTGGCTTGTAGGATATATAAAGTAAAGACACCTGTATAACAGGAGACA 2340
Qy 2341 TTAGATCAAAACCTGAGTCCACCAATTTTCCCAAAACAGAGATTTTGTCAAGAGAG 2400
Db 2341 TTAGATCAAAACCTGAGTCCACCAATTTTCCCAAAACAGAGATTTTGTCAAGAGAG 2400
Qy 2401 ACTACCAAAAGGAGGCTGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 2460
Db 2401 ACTACCAAAAGGAGGCTGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 2460
Qy 2461 ATATCTAAGAGACTGTTTAA 2481
Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 3

BC052433

LOCUS

DEFINITION

MUS musculus RIKEN cDNA 1300006C19 gene, mRNA (cdna clone MSC:64679)

IMAGE:6837097, complete cds.

ACCESSION

BC052433

VERSION

BC052433.1

KEYWORDS

MGC.

SOURCE

MUS musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 4236)

Strauberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, S.I., Wang, J., Hsieh, F.,

Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Sheezy, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Lequellano, N.A., Peters, G.J.,

Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Buetow, K.H., Krzywicki, M.I., Skalska, U., Smal, D.E.,

ORIGIN

Query Match 100.0%; Score 2481; DB 6; Length 2481;
Best Local Similarity 100.0%; Pred No. 0;
Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGAGCCCTCGGCCCCGGAGAGCAAGTGTGCTCTCAATCTGTCCTCCCGTGG 60
DB 1 ATGGCGGAGCCCTCGGCCCCGGAGAGCAAGTGTGCTCTCAATCTGTCCTCCCGTGG 60

QY 61 AGTGGCTCATGGCCCTGGGAAACAGCCGCGCCAGCCACACGCGGCCCGGCGCCAGTGC 120
DB 61 AGTGGCTCATGGCCCTGGGAAACAGCCGCGCCAGCCACACGCGGCCCGGCGCCAGTGC 120

QY 121 GCGCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 121 GCGCACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180

QY 181 GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 181 GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

QY 241 TGGCTTGGCGGCTTCAGCTCGGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 300
DB 241 TGGCTTGGCGGCTTCAGCTCGGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 300

QY 301 GAGTTCGACCGGCTGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCTAT 360
DB 301 GAGTTCGACCGGCTGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGGTTCTAT 360

QY 361 GAATTTTAAATGGTTTGAAGAGAGATGGTATCCACTAGGAAGATAGTAGGGT 420
DB 361 GAATTTTAAATGGTTTGAAGAGAGATGGTATCCACTAGGAAGATAGTAGGGT 420

QY 421 ACTGTTTACCAGGGTGGATGAACCGCTGGGCTTATTCAATGGATTTTAAATACATG 480
DB 421 ACTGTTTACCAGGGTGGATGAACCGCTGGGCTTATTCAATGGATTTTAAATACATG 480

QY 481 AACATACTGTACATAAGAGAGATGATGTTGCTTGCACCACTTTTACGCGCTT 540
DB 481 AACATACTGTACATAAGAGAGATGATGTTGCTTGCACCACTTTTACGCGCTT 540

QY 541 ACATCTATATCTCTTCTCTTCAAGAGAACTTTGGAACCAAGAGCAGGACTTTTA 600
DB 541 ACATCTATATCTCTTCTCTTCAAGAGAACTTTGGAACCAAGAGCAGGACTTTTA 600

QY 601 GCTGCTTGTGTTTATGCTATTTGACAGGCTACATATCTCGGTGCTGATGCTGATCT 660
DB 601 GCTGCTTGTGTTTATGCTATTTGACAGGCTACATATCTCGGTGCTGATGCTGATCT 660

QY 661 GATAATGAAGCATTTGCTATTTTTCAGTTTCACTTCACTATCTATTTTATGGGTAAATCT 720
DB 661 GATAATGAAGCATTTGCTATTTTTCAGTTTCACTTCACTATCTATTTTATGGGTAAATCT 720

QY 721 GTAAATCTGGGTGCTGTTTGGACAAATGCTGCTGCTTATCTTATTTCTATATGCTC 780
DB 721 GTAAATCTGGGTGCTGTTTGGACAAATGCTGCTGCTTATCTTATTTCTATATGCTC 780

QY 781 TCTGCTTGGGTGCTGTTTATGATTTTATCATCAATCTTATTCAGTGTATTTTGTGTTG 840
DB 781 TCTGCTTGGGTGCTGTTTATGATTTTATCATCAATCTTATTCAGTGTATTTTGTGTTG 840

QY 841 TTACTGTAGCAGATACAGCAAGAGTCTACATAGCATATAGCATTTTCTTACATGTTG 900
DB 841 TTACTGTAGCAGATACAGCAAGAGTCTACATAGCATATAGCATTTTCTTACATGTTG 900

QY 901 GGTTTAAATTTTATCAATGACATACCTTTTGTGGATTTCCAGCAATCAGAACAGTGAA 960
DB 901 GGTTTAAATTTTATCAATGACATACCTTTTGTGGATTTCCAGCAATCAGAACAGTGAA 960

QY 961 CACATGGAGCTGCGAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020
DB 961 CACATGGAGCTGCGAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020

QY 1021 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1080
DB 1021 AGAGACCGATTAAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGTATCACTAGCT 1080

QY 1081 GCAGGTGCTGTGTTCTTGTAGTGTATCTATTTGACTTATACAGTTTACATTGCGACCATGG 1140
DB 1081 GCAGGTGCTGTGTTCTTGTAGTGTATCTATTTGACTTATACAGTTTACATTGCGACCATGG 1140

QY 1141 AGTGGCAGGTTTTTATTCATTGTGGGATCTGGGTATGCAAAAATACACATTCCAATTATT 1200
DB 1141 AGTGGCAGGTTTTTATTCATTGTGGGATCTGGGTATGCAAAAATACACATTCCAATTATT 1200

QY 1201 GCATCAGTGTCTGAGCATCAACCTAGCCTGGGTGCTCTTCTTCTTGTATCTACATATT 1260
DB 1201 GCATCAGTGTCTGAGCATCAACCTAGCCTGGGTGCTCTTCTTCTTGTATCTACATATT 1260

QY 1261 CTTGTATGTACCTTCCAGCAGGCGCTTTGGTCTCTCATCAAAAATATCAACGATGAAAGA 1320
DB 1261 CTTGTATGTACCTTCCAGCAGGCGCTTTGGTCTCTCATCAAAAATATCAACGATGAAAGA 1320

QY 1321 GTATTTTGTGCTTATATGCAATCAGTGTCTCTACTTTGCTGGAGTGTGTCGACTG 1380
DB 1321 GTATTTTGTGCTTATATGCAATCAGTGTCTCTACTTTGCTGGAGTGTGTCGACTG 1380

QY 1381 ATGTTGACTTTGACTCGTGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTTTT 1440
DB 1381 ATGTTGACTTTGACTCGTGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTTTT 1440

QY 1441 GAGCATTTTGGGGATGACATGAAAGGGAATCCACCTGTGGAGCAGCAGTGTAT 1500
DB 1441 GAGCATTTTGGGGATGACATGAAAGGGAATCCACCTGTGGAGCAGCAGTGTAT 1500

QY 1501 GAGGATGACAAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTGAGGAAACAT 1560
DB 1501 GAGGATGACAAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTGAGGAAACAT 1560

QY 1561 GCAACTGACAGGAAAGAACTGAAGGGGATTTAGGCCCTTAATATAAAGAGCATTTCTACC 1620
DB 1561 GCAACTGACAGGAAAGAACTGAAGGGGATTTAGGCCCTTAATATAAAGAGCATTTCTACC 1620

QY 1621 ATGTTGATGCTGATGCTATTGATGATGTTTGTCTGCTCACTGTACCTGGGTCAACAGCAAT 1680
DB 1621 ATGTTGATGCTGATGCTATTGATGATGTTTGTCTGCTCACTGTACCTGGGTCAACAGCAAT 1680

QY 1681 GCCTCTCTAGTCCAAAGTGTGCTGCGCTCATACAAATCATGATGGCAGCAGGAATATC 1740
DB 1681 GCCTCTCTAGTCCAAAGTGTGCTGCGCTCATACAAATCATGATGGCAGCAGGAATATC 1740

QY 1741 TTAGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAAATACAGATGAACATGCAAGA 1800
DB 1741 TTAGATGATTTTAGAAGCTTACTTTTGGCTAAGGCAAAAATACAGATGAACATGCAAGA 1800

QY 1801 GTATGCTCTGGGTGATTTATGCTATCAGATAGTGGATGCTTAATAGACTACGTTG 1860
DB 1801 GTATGCTCTGGGTGATTTATGCTATCAGATAGTGGATGCTTAATAGACTACGTTG 1860

QY 1861 GTGGATAATTAACACCTCGGAATTAACCCACATAGCAGTGGTGGGAAAAGCTATGCTTCT 1920
DB 1861 GTGGATAATTAACACCTCGGAATTAACCCACATAGCAGTGGTGGGAAAAGCTATGCTTCT 1920

QY 1921 AATGAACACAGCAGCCTTAAATCATAGGACTCTAGATGTAGATATGTTTGGTATT 1980
DB 1921 AATGAACACAGCAGCCTTAAATCATAGGACTCTAGATGTAGATATGTTTGGTATT 1980

QY 1981 TTTGGAGGGGTTATTGGCTATTCTGGTGTATGATATCAACAAAATTTCTCTGGATGTTAGG 2040
DB 1981 TTTGGAGGGGTTATTGGCTATTCTGGTGTATGATATCAACAAAATTTCTCTGGATGTTAGG 2040

QY 2041 ATAGCTCAAGGAGAAATCCCAAGACATTCGGGAAGTGTATTTTACCCACAGGGA 2100
DB 2041 ATAGCTCAAGGAGAAATCCCAAGACATTCGGGAAGTGTATTTTACCCACAGGGA 2100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 22:52:10 ; Search time 10321.1 Seconds
(without alignments)
11367.538 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcgagccctcgccccc.....tatctaagaagactgttttaa 2481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 2364484745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hlg.*

3: gb_in.*

4: gb_cm.*

5: gb_ov.*

6: gb_pet.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 2481 | 100.0 | 2481 | 6 | AX799082 Sequence |
| 2 | 2481 | 100.0 | 2481 | 9 | AY074880 Homo sapi |
| 3 | 2081.8 | 83.9 | 4236 | 10 | BC052433 Mus muscu |
| 4 | 2078.6 | 83.8 | 2710 | 6 | AX799084 Sequence |
| 5 | 1328.8 | 53.6 | 1664 | 6 | CQ782554 Sequence |
| 6 | 1328.8 | 53.6 | 1664 | 6 | BD127193 Sequence |
| 7 | 1328.8 | 53.6 | 1664 | 9 | AK074587 Homo sapi |
| 8 | 1307.4 | 52.7 | 1734 | 10 | BC003206 Mus muscu |
| 9 | 972 | 39.2 | 2546 | 6 | AX099510 Sequence |
| 10 | 972 | 39.2 | 2546 | 6 | BD063986 Secreted |
| 11 | 955.8 | 38.5 | 2417 | 3 | AF132552 Drosophil |
| 12 | 955.8 | 38.5 | 2417 | 6 | AX799088 Sequence |
| 13 | 955.8 | 38.5 | 2699 | 6 | CQ589353 Sequence |
| 14 | 954 | 38.5 | 2510 | 6 | AX136135 Sequence |
| 15 | 954 | 38.5 | 2510 | 6 | BD123520 Homo sapi |
| 16 | 954 | 38.5 | 2510 | 9 | AK075380 Homo sapi |
| 17 | 923.2 | 37.2 | 2503 | 6 | AX017997 Sequence |
| 18 | 923.2 | 37.2 | 2503 | 6 | BD137303 Human nuc |
| 19 | 861.4 | 34.7 | 2284 | 6 | AX882932 Sequence |

| | | | | | |
|----|-------|------|--------|----|--------------------|
| 20 | 861.4 | 34.7 | 2284 | 6 | BD160013 |
| 21 | 861.4 | 34.7 | 2284 | 9 | AK027789 Homo sapi |
| 22 | 833.4 | 33.6 | 2620 | 5 | CR386955 Gallus ga |
| 23 | 761.6 | 30.7 | 4522 | 6 | CQ589352 Sequence |
| 24 | 761.6 | 30.7 | 162921 | 3 | AC007853 Drosophil |
| 25 | 761.6 | 30.7 | 181132 | 3 | AC008206 Drosophil |
| 26 | 761.6 | 30.7 | 227219 | 3 | AE003750 Drosophil |
| 27 | 754.4 | 30.4 | 957 | 9 | BC015880 Homo sapi |
| 28 | 753.2 | 30.4 | 2508 | 10 | BC013054 Mus muscu |
| 29 | 749.6 | 30.2 | 75550 | 2 | AC018145 Drosophil |
| 30 | 712.8 | 28.7 | 787 | 6 | AX869456 Sequence |
| 31 | 712.8 | 28.7 | 787 | 6 | BD149518 Primer fo |
| 32 | 695.8 | 28.0 | 2481 | 5 | BC067313 Xenopus t |
| 33 | 689.2 | 27.8 | 22566 | 6 | AX489023 Sequence |
| 34 | 688.8 | 27.8 | 2555 | 8 | BT012881 Lycopersi |
| 35 | 671 | 27.0 | 764 | 6 | AX136480 Sequence |
| 36 | 671 | 27.0 | 764 | 6 | BD123720 Secretory |
| 37 | 670.6 | 27.0 | 2736 | 10 | BC037612 Mus muscu |
| 38 | 670.6 | 27.0 | 3094 | 6 | AX799090 Sequence |
| 39 | 670.6 | 27.0 | 3094 | 10 | MUSTRPR |
| 40 | 667.4 | 26.9 | 3666 | 10 | AX129027 Mus muscu |
| 41 | 655 | 26.4 | 2855 | 6 | CQ573675 Sequence |
| 42 | 646.6 | 26.1 | 2118 | 9 | BT007100 Homo sapi |
| 43 | 646.6 | 26.1 | 2118 | 12 | BT008132 Synthetic |
| 44 | 646.6 | 26.1 | 2293 | 9 | HUNTMC |
| 45 | 646.6 | 26.1 | 2450 | 9 | EC048348 Homo sapi |

ALIGNMENTS

RESULT 1
AX799082
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Sequence 1 from Patent WO03054008.
AX799082
AX799082.1 GI:37605057
Homo sapiens (human)
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Ferreault,C. and McBride,K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 1 03-JUL-2003;
Compatisgene Inc. (CA)

FEATURES
source
Location/Qualifiers
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2481
/codon_start=1
/product="SIMP"
/protein_id="CAE48447.1"
/db_xref="GI:37605057"
/translation="MAESAPESKHSLNSPWSGLMALGNRHHGHGGAQCAHKA
AGAAPAPAGLGGSLGSPAGWQSLSTFLFLAWLAGSRFLFAVIRFESITIEHF
DPFNRYRTHLASHGFYFLNWFDERAWPLRVGGTVPGMLITAGLHWILNTL
NITHIRVVCVFLAPTFESGLTSISTFLTRELNQGAGLAAACFATVPGYISRSVAG
SPDNIGIAIFALQFTYLVKSVKTVFVMTCCLSYFVYVAGAGYVFIINLPIH
FVLLLMQRYKRVYAVSTYIVGLILSMQIPFVGQPIRTSEHMAAGVFLALQAY
AFVLLYRLTKQSFOTLFFGLVSLAAGAVFLVILTYTGYTAPWSGRVSLWDYGY
AKIHPIIASVSEHQPTWVSFFDLHLVCTPAGLWFCIKININDEVRVAYISA
VYFAGVWFLMLTILTPVVCMSALAFNVEHILGDDMKENPVEDSDEDDKRNQG
NLYDAGKVRKHATEQETEGGLGNKISVITMLMLMLLMFAVCTWVTSYSSPS
VWNSYHDGTGNILDDFREYFWLRQNTDEHARVMSWVDYGYQAGMARFTLVNN
TWNNSHIALVGKMSNETAAYKIMRLDVIDVLIYFVGIVGSGDDINKFLMWVRIA
EGEHPKIDRESGYPTQGEFRVDRAGSPTLNCILMYKMSYRFGEGQLDPTPGDR
TNNAGINKDKFKHLEAFTSEHVLVRIYKVKAPDNRETLDHKPRVTNIFPQKYL
KXTTKRGYIKNKLIFKKGKISKKTV"

CDS

[illegible]

| | | |
|------|--|------|
| 1559 | ATGCAACTGGAACAGGAAAAAACTGAGAGAGGATTAGGCCCTTAATATAAAAGCAATTGTCA | 1618 |
| Qy | | |
| 1491 | GTACTTAAGCATGATGCCACAGCAGTACTGGCGTCAAGCTTCAACCTGAAAGTAGTATTGTTA | 1550 |
| Db | | |
| 1619 | CCATGCTGATGCTCATGCTATTGATGATGTTTGTGCTCCACTGTACTCTGGGTCAACAAGCA | 1678 |
| Qy | | |
| 1551 | TTTTGGCGGTTCTAATGCTGTTGATGATGTTCCGTGTCACCTGACAGTGGGTGACACAGCA | 1610 |
| Db | | |
| 1679 | ATGCTCTACTCTPAGTCCAAAGTGTAATGCTCTGGCCTCATACAA---TCATGATGGCAACAGGA | 1735 |
| Qy | | |
| 1611 | ATGCCTACTCCAGTCCCTCCATTGTTCTTGCTTTGCCATTCCACAACAGTCAAGATGGATCCCGCA | 1670 |
| Db | | |
| 1736 | ATATCTTATAGATGATTTTATAGAGAAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATG | 1795 |
| Qy | | |
| 1671 | ACATTTTATAGACATTTTACAGAGAGCTTACTATGCTGGCTTTCGACGAACACTCTCCGATGATG | 1730 |
| Db | | |
| 1796 | CACGAGTAAATGCTCTGGTGGGATTATGGCTATCAGATAGCTGGAAATGCTTAATAGAACTA | 1855 |
| Qy | | |
| 1731 | CTCGCGTTATGCTCTGGTGGGATTACGGATACCGAGATAGCGGAATGCGAAACAGAACGA | 1790 |
| Db | | |
| 1856 | CGTTGGTGGATAATAACACTGGGAATAACAGCCACATAGCACTGGTGGGAAAAGCTATGT | 1915 |
| Qy | | |
| 1791 | CCGTAGTGGATAATAATACGTGGAAACAATAGTCAATAGCGCTGGTTGGCAAGGCAATGT | 1850 |
| Db | | |
| 1916 | CTTCTAATGAACACAGCAGCCTATAAAATCATGAGGACTCTAGATGTAGATTATGTTTGG | 1975 |
| Qy | | |
| 1851 | CTTCAACCGAGGAGAGTCTACGAAATATGACATCTCTTGACGTGGACTACGTTTGG | 1910 |
| Db | | |
| 1976 | TTATTTTTGGAGGGGTTATTTGGCTATCTGGTGAATGATATCAACAAATTTCTCTGGATGG | 2035 |
| Qy | | |
| 1911 | TGATCTTTGGCGGGTGTGATCGGCTATTCTGGCGATATCAACAAGTTCTCTGTGGATGG | 1970 |
| Db | | |
| 2036 | TTAGATAGCTGAAGGAGAAACATCCCAAAGACATTCGGAAAGTCACTATTTTACCCAC | 2095 |
| Qy | | |
| 1971 | TCCGAATTTGCTGAGGAGAGACATCCCAAGSACATTAAGSAAAGCATTTTACCAGACC | 2030 |
| Db | | |
| 2096 | AGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACTTTGTTGAATTCCTTATGTATATA | 2155 |
| Qy | | |
| 2031 | CGCGTGAATTCAGGGTAGATGCCGAAGGTGCTCCGGCCCTGCTCAACTGCGCTTAATGTACA | 2090 |
| Db | | |
| 2156 | AAATGTCATCTACAGATTTGGAGAAATGCAGCTGGATTTCGTACACCCCCAGGTTTG | 2215 |
| Qy | | |
| 2091 | AATTAAGCTACTACAGATTCGGGGAATTGAAGTTGGACTACAGAGGTCCTCTGATATG | 2150 |
| Db | | |
| 2216 | ACCGAAACGTAATGCTGAGATGGGAAATAAGSACATTAATTTCAAACATTTGGAGAAG | 2275 |
| Qy | | |
| 2151 | ATCGCACACGTAACGCGGCTCAITTTGGGAATAAGSACTTCGATCTGACCTACCTGGAGAGG | 2210 |
| Db | | |
| 2276 | CGTTTACATCAGAACACTGGCTGTTTAGGATATATAAAGTAAAGCACTCATATAACAGG | 2335 |
| Qy | | |
| 2211 | CTCTACCAACAGAACACTGGCTTGTTCGATCTTATAGGGTGAAGAGCGCATGAGTTCA | 2270 |
| Db | | |
| 2336 | AGACATTAGATCAAAAACCTCGAGTCAACCAATTTTCCAAAAACAGAAGTATTGTCAA | 2395 |
| Qy | | |
| 2271 | ATAGACCATCACTGAAGACCAAGGAGAGAACGATTCTCCAGCAAACTTCATTTTCGAGAA | 2330 |
| Db | | |
| 2396 | AGAAGACTACCAAAGGAAGCGGTGCTACATTAAATAATTAAGTCGTTTTTAAGAAAGCA | 2455 |
| Qy | | |
| 2331 | AGNACTTAAGGTCGCAAG---GGCTACATACGAAACCGAGTCGGTTGTTTAAGGAA | 2397 |
| Db | | |
| 2456 | AGAAAAATATCTAAGAAGAC | 2474 |
| Qy | | |
| 2388 | AACGAACCTTGAATAAAC | 2406 |
| Db | | |

| | |
|-----------|-----------------------------------|
| RESULT 15 | |
| AAF93772 | |
| ID | AAF93772 standard; cDNA; 2510 BP. |
| XX | |
| XX | AAF93772; |
| XX | |
| XX | |
| DT | 23-MAY-2001 (first entry) |
| XX | |

| | |
|-----------|---|
| RESULT 14 | |
| ABU13247 | |
| ID | ABL13247 standard; cDNA; 2699 BP. |
| XX | |
| XX | ABL13247; |
| DT | 26-MAR-2002 (first entry) |
| XX | |
| DE | Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223. |
| XX | |
| KW | Drosophila; developmental biology; cell signalling; insecticide; |
| KW | pharmaceutical; gene; ss. |
| XX | |
| OS | Drosophila melanogaster. |
| XX | |
| PN | WO200171042-A2. |
| XX | |
| PD | 27-SEP-2001. |
| XX | |
| PF | 23-MAR-2001; 2001WO-US009231. |
| XX | |
| PR | 23-MAR-2000; 2000US-0191637P. |
| PPR | 11-JUL-2000; 2000US-00614150. |
| XX | |
| PA | (PEKE) PE CORP NY. |
| XX | |
| PI | Venter JC, Adams M, Li PWD, Myers EW; |
| XX | |
| PFI | WPI; 2001-656860/75. |
| DR | P-P9DB; ABB69144.. |
| DR | |
| PT | New isolated nucleic acid detection reagent for detecting 1000 or more |
| PXT | genes from Drosophila and for elucidating cell signaling and cell-cell |
| PT | interactions. |
| XX | |
| XX | Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English. |
| XX | |
| CC | The invention relates to an isolated nucleic acid detection reagent |
| CC | capable of detecting 1000 or more genes from Drosophila. The invention is |
| CC | useful in developmental biology and in elucidating cell signalling and |
| CC | cell-cell interactions in higher eukaryotes for the development of |
| CC | insecticides, therapeutics and pharmaceutical drugs. The invention |
| CC | discloses genomic DNA sequences (ABU16176-ABU130511), expressed DNA |
| CC | sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737- |
| CC | ABB72072). The sequence data for this patent did not form part of the |
| CC | printed specification, but was obtained in electronic format directly |
| CC | from WIPO at ftp.wipo.int/pub/published_pct_sequences |
| XX | |
| SQ | Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other; |
| | |
| | Query Match 38.58; Score 955.8; DB 4; Length 2699; |
| | Best Local Similarity 64.8%; Pred. No. 2.9e-236; |
| | Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4 |
| QY | 191 AGCGGCTGGGTGCAGTCGCTTCTCCTCCATCATCTCTCTCTGCCCTGCTGCGG 250 |
| Db | 111 AGTGTTGGCTACAGCAGCCTAATCACCTTGCCATCCTGCTAAATCGCTGCTGGCGG 170 |
| QY | 251 GCTTCAGTCGCGCTCTTCGCGCTCATCGCTTCGAAGAGCATATCCACGAGTTCGACC 310 |
| Db | 171 GATTTCCTCTCGGCTCTTCGCGCTCATCGTTTTGATGTCGATTAATCATGATTTGATC 230 |
| QY | 311 CGTGGTTAACTATAGATCAACACATCATCTTGCATCTCATCGGTTCTATGAATTTTAA 370 |
| Db | 231 CGTGGTTCAACTACCGGGCCACCGCCTACATGCTGCAGAATGTTGGTACAACCTTCCTCA 290 |
| QY | 371 ATTGGTTTGATGAAGAGCATGTATTCCTACTAGGAGAGATAGTAGTGTGTTTACC 430 |
| Db | 291 ACTGTTTCGACGACGCGCGATGATACCGCTCGCAGGANTTGGGCGGTCACGTCATC 350 |
| QY | 431 CAGGGTTGATGAACCGCTGGCCTTATTCAATTGGATTTTAAATACATTGAACATACTG 490 |
| Db | 351 CCGCCCTGATGATTAATGCTCCGCGCGGAATCCATTGGCTGTCGACGCTACTCAACATACGG 410 |

QY 2410 AGAAGCGTGGCTACATTAATAAGCTGTTTAAAGAGGCAAGAAATATCTAAG 2469
 DB 901 AGGAGCGTGGCTACATTAATAAGCTGTTTAAAGAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972

RESULT 12
 AAF98463
 ID AAF98463 standard; cDNA; 2546 BP.
 XX
 AC AAF98463;
 DT 07-JUN-2001 (first entry)
 XX
 DE Human cDNA clone CT585_1 sequence 8BQ ID 150.
 XX
 KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
 KW differentiation; immune system modulator; tissue growth; chemotactic;
 KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
 KW haematopoiesis.
 XX
 OS Homo sapiens.
 XX
 PN WO200119988-A1.
 XX
 PD 22-MAR-2001.
 XX
 PF 14-SEP-2000; 2000WO-US025135.
 XX
 PR 17-SEP-1999; 99US-00398829.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
 XX
 DR WPI; 2001-244801/25.
 XX
 DR P-PSDB; AAB90727.
 XX
 PT Isolated nucleic acids encoding polypeptides, useful for modulating e.g.
 PT cytokine and cell proliferation/differentiation activity, the immune
 PT system and hematopoiesis regulating activity.
 XX
 PS Disclosure; Page 476-477; 557pp; English.
 XX
 CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
 CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
 CC tissue types, and may be used in the prevention, treatment and diagnosis
 CC of diseases associated with inappropriate protein expression. The
 CC polypeptides and nucleic acids may be used as nutrients or to modulate
 CC cytokine and cell proliferation/differentiation activity and may also be
 CC involved in modulation of the immune system. The cDNA sequences,
 CC proteins, their agonists and/or antagonists exhibit haematopoiesis
 CC regulating activity; tissue growth activity; activin/inhibin activity;
 CC chemotactic/chemokinetic activity; haemostatic and thrombolytic activity;
 CC receptor/ligand activity; anti-inflammatory activity; haematopoiesis
 CC activity; cadherin/tumour suppressor activity; and/or tumour inhibition
 CC activity. Included in the invention are probes represented in AAF98490 -
 CC AAF98572 which are specific for the cDNA clones encoding the secreted
 CC proteins
 XX
 SQ Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;
 39, 28; Score 972; DB 5; Length 2546;
 Query Match
 Best Local Similarity 100.0%; Pred. NO. 1.7e-240;
 Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 AAAAGAAACCAAGGAAATTTGTATGATGAGCGAGGTAAGTGAAGGAAACATCAACTGAA 1569
 |||||

DB 1 AAAAGAAACCAAGGAAATTTGTATGATGAGCGAGGTAAGTGAAGGAAACATCAACTGAA 60
 QY 1570 CAGAAAAAAGCTGAGAGGATTTAGCCCTAATAATAAAAGCATTTGTCCACCATTTGATG 1629
 DB 61 CAGAAAAAAGCTGAGAGGATTTAGCCCTAATAATAAAAGCATTTGTCCACCATTTGATG 120
 QY 1630 CTGATCTATTGATGATGTTTCTGCTCACTGTACTGGGTCAAGCAATATGCTTACTCT 1689
 DB 121 CTGATCTATTGATGATGTTTCTGCTCACTGTACTGGGTCAAGCAATATGCTTACTCT 180
 QY 1690 AGTCCAAAGTGTAGTCTGCTCTCATCAATATGATGACCCAGCAATATCTTATGATGAT 1749
 DB 181 AGTCCAAAGTGTAGTCTGCTCTCATCAATATGATGACCCAGCAATATCTTATGATGAT 240
 QY 1750 TTTAGAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGACGAGTAAATGCT 1809
 DB 241 TTTAGAGAAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAACATGACGAGTAAATGCT 300
 QY 1810 TGGTGGATTTATGGCTTATCAGATAGCTGGAATGGCTAATAAGAACTACCTTGGTGATAAT 1869
 DB 301 TGGTGGATTTATGGCTTATCAGATAGCTGGAATGGCTAATAAGAACTACCTTGGTGATAAT 360
 QY 1870 AACACCTGGAATAACAGCCACATAGCACTGCTGGGAAAGCTATGCTTCTTAATGAACA 1929
 DB 361 AACACCTGGAATAACAGCCACATAGCACTGCTGGGAAAGCTATGCTTCTTAATGAACA 420
 QY 1930 GCAGCCTATAAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATTTTGGAGGG 1989
 DB 421 GCAGCCTATAAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATTCTGGTGATGATATCAAAATTTCTCTGGATGGTTAGGATAGCTGAA 2049
 DB 481 GTTATTGGCTATTCTGGTGATGATATCAAAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGAGAAATCCGT 2109
 DB 541 GGAGAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCACAGGAGAAATCCGT 600
 QY 2110 GTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCGCTTATGATATAAAATGTCATCTAC 2169
 DB 601 GTAGACAAAGCAGGATCCCTTACTTTGTTGAATTCGCTTATGATATAAAATGTCATCTAC 660
 QY 2170 AGATTGGGAGAAATGAGCTGGATTTTGTACACCCCGAGGTTTGTACCGAACACGTAAT 2229
 DB 661 AGATTGGGAGAAATGAGCTGGATTTTGTACACCCCGAGGTTTGTACCGAACACGTAAT 720
 QY 2230 GCTGAGATTGGAAATAAGGACATTAATTCAAACATTTTGGAGAGAGCCTTTTACATCAGAA 2289
 DB 721 GCTGAGATTGGAAATAAGGACATTAATTCAAACATTTTGGAGAGAGCCTTTTACATCAGAA 780
 QY 2290 CACTGCTTGTAGGATATATAAGTAAAGCACTGATTAACAGGGAGACATTTAGATCAC 2349
 DB 781 CACTGCTTGTAGGATATATAAGTAAAGCACTGATTAACAGGGAGACATTTAGATCAC 840
 QY 2350 AAACCTCGAGTCACCAACATTTTCCCAAAAGAGAGTATTTGTCAAAGAGACTACCAA 2409
 DB 841 AAACCTCGAGTCACCAACATTTTCCCAAAAGAGAGTATTTGTCAAAGAGACTACCAA 900
 QY 2410 AGGAGCGTGGCTACATTAATAAGCTGTTTAAAGAGGCAAGAAATATCTAAG 2469
 DB 901 AGGAGCGTGGCTACATTAATAAGCTGTTTAAAGAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTGTTTAA 2481
 DB 961 AAGACTGTTTAA 972

RESULT 13
 ADD94789
 ID ADD94789 standard; DNA; 2417 BP.
 XX
 XX ADD94789;
 XX

Db 661 TTTTGTGGATCCAGCAATCAGAAACAGTGAACATGGCAGCTGCAGGTGTCTTTC 720
 QY 987 ATTGCTGCAAGCTTATCTTCTTCTGAGTATCTGAGAGCCGATTAACAAAACAGAGT 1046
 Db 721 ATTGCTGCAAGCTTATCTTCTTCTGAGTATCTGAGAGCCGATTAACAAAACAGAGT 780
 QY 1047 CCAGACCCCTTTCTTTTGGGTGTATCAGTGTGAGTGTCTGCTTCTTCTTCTTCTTCT 1106
 Db 781 CCAGACCCCTTTCTTTTGGGTGTATCAGTGTGAGTGTCTGCTTCTTCTTCTTCTTCT 840
 QY 1107 CTATTTGACATATACAGGTTACATTCACCATGGAGTGGCAGTTTATTTCTTCTTCTTCT 1166
 Db 841 CTATTTGACATATACAGGTTACATTCACCATGGAGTGGCAGTTTATTTCTTCTTCTTCT 900
 QY 1167 TACTGGGTATGCAAAATACACATTCACATTTATGATGATGATGATGATGATGATGATGAT 1226
 Db 901 TACTGGGTATGCAAAATACACATTCACATTTCAATTTATGATGATGATGATGATGATGAT 960
 QY 1227 GACTTGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1286
 Db 961 GACTTGGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020
 QY 1287 TTGTTCTGATCAAAATATCAACGATGAAGAGATTTTGT 1329
 Db 1021 TTGTTCTGATCAAAATATCAACGATGAAGAGATTTTGT 1063

RESULT 11

RAV44866
 ID AAV44866 standard; cDNA; 2546 BP.

XX AC AAV44866;

XX DT 21-OCT-1998 (first entry)

XX DE Clone CT585_1 coding sequence.

XX KW Secreted protein; nutritional source; cell proliferation activity;
 KW cell differentiation activity; immune stimulant; tissue growth activator;
 KW haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
 KW tumour inhibitor; clone CT585_1; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX CDS 112..972
 XX FT /*tag= a

XX FN WO9825962-A2.

XX PD 18-JUN-1998.

XX PF 12-DEC-1997; 97WO-US023224.

XX PR 13-DEC-1996; 96US-00766263.

XX PR 11-DEC-1997; 97US-00989232.

XX PA (GEM) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
 PI Spaulding V, Agostino M;

XX DR WPI; 1998-362424/31.

XX DR P-PSDB; AAW69247.

XX PT New isolated polynucleotides - obtained from human adult testis, human
 PT adult ovary, human adult brain and human adult heart cDNA libraries.

XX PS Claim 35; Page 79-81; 108pp; English.

XX CC This sequence represents a polynucleotide of the invention, and encodes a
 CC secreted protein. It was isolated from a human adult brain cDNA library,
 CC and is designated clone CT585_1. The DNA sequences and encoded

CC polypeptides can be used as nutritional sources or supplements, or may
 CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
 CC immune stimulating or suppressing activity, haematopoiesis regulating
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC activin/inhibin activity, chemostatic/chemokinetic activity,
 CC cadherin/tumour invasion suppressor activity, tissue growth activity,
 CC tumour inhibition activity or other activities

XX SQ Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;

Query Match 39.2%; Score 972; DB 2; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 1.7e-240;

Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 AAAAGAAACCAAGGAAATTTGTATGATAGGAGGATTAAGTGAGGAAACATGCAACTGAA 1569

Db 1 AAAAGAAACCAAGGAAATTTGTATGATAGGAGGATTAAGTGAGGAAACATGCAACTGAA 60

QY 1570 CAGGAAACCAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGAT 1629

Db 61 CAGGAAACCAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGATTAAGGAGGAT 120

QY 1630 CTGATGCTATTCATGATGTTTGTCTCTCCACTGTACCTGGGTCAACAAGCAATGCTACTCT 1689

Db 121 CTGATGCTATTCATGATGTTTGTCTCTCCACTGTACCTGGGTCAACAAGCAATGCTACTCT 180

QY 1690 AGTCCAAAGTGTAGTCTCTGSCCTCATACATCATGATGSCCAGGAAATATCTTAGATGAT 1749

Db 181 AGTCCAAAGTGTAGTCTCTGSCCTCATACATCATGATGSCCAGGAAATATCTTAGATGAT 240

QY 1750 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAAGTAATGTCT 1809

Db 241 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAAGTAATGTCT 300

QY 1810 TGTGGGATTATGGCTATCAGATAGCTGAATGCTAATGAATGCTGTTGGTGAATTAAT 1869

Db 301 TGTGGGATTATGGCTATCAGATAGCTGAATGCTAATGAATGCTGTTGGTGAATTAAT 360

QY 1870 AACACCTGGAATTAACAGCCACATAGCATGTGTGGGAAAGCTATGTCTTTAATGAACA 1929

Db 361 AACACCTGGAATTAACAGCCACATAGCATGTGTGGGAAAGCTATGTCTTTAATGAACA 420

QY 1930 GCAGCCTATAAATCATGAGGACTCTAGATGATGATTAATGCTGTTGGTGAATTAAT 1989

Db 421 GCAGCCTATAAATCATGAGGACTCTAGATGATGATTAATGCTGTTGGTGAATTAAT 480

QY 1990 GTTATTGGCTATTCTGTGATGATTAACAATAATTTCTGGATGTTAGGATAGCTGAA 2049

Db 481 GTTATTGGCTATTCTGTGATGATTAACAATAATTTCTGGATGTTAGGATAGCTGAA 540

QY 2050 GGAGAACATCCCAAGACATTCGGGAAAGTCACTATTTTACCCACAGGAGGATTCCT 2109

Db 541 GGAGAACATCCCAAGACATTCGGGAAAGTCACTATTTTACCCACAGGAGGATTCCT 600

QY 2110 GTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATAAATGTCTACTAC 2169

Db 601 GTAGACAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGTATAAATGTCTACTAC 660

QY 2170 AGATTGTGAGGAATGACGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2229

Db 661 AGATTGTGAGGAATGACGTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 720

QY 2230 GCTGAGATTGGAATTAAGGACATTAATTTCAACATTTGGAAGAAGCTTTTACATCAGAA 2289

Db 721 GCTGAGATTGGAATTAAGGACATTAATTTCAACATTTGGAAGAAGCTTTTACATCAGAA 780

QY 2290 CACTGCTGTGTAGGATATATAAGTAAAGCCTGATACAGGAGACATTAGATCAC 2349

Db 781 CACTGCTGTGTAGGATATATAAGTAAAGCCTGATACAGGAGACATTAGATCAC 840

QY 2350 AAACCTCGAGTCACCAACATTTTCCAAAACAGAGTATTTGTCAAAAGAGACTACAAA 2409

Db 841 AAACCTCGAGTCACCAACATTTTCCAAAACAGAGTATTTGTCAAAAGAGACTACAAA 900

QY 1227 GACTTGGGTGCTTCTTCTTTGATCTACATATTTCTTGATGACCTCCAGAGGCT 1286
 DB 961 GACTTGGGTGCTTCTTCTTTGATCTACATATTTCTTGATGACCTCCAGAGGCT 1020
 QY 1287 TTGGTCTGCAATCAAAATATCAACGATGAAAGATTTGTT 1329
 DB 1021 TTGGTCTGCAATCAAAATATCAACGATGAAAGATTTGTT 1063

RESULT 10

ADN60663

ID ADN60663 standard; cDNA; 1543 BP.

XX AC ADN60663;

XX DT 01-JUL-2004 (first entry)

XX DE Human secreted polynucleotide #123.

XX KW Human; secreted polynucleotide; gene; ss; autoimmune disease;
 KW rheumatoid arthritis; hyperproliferative disorder; neoplasm; breast;
 KW liver; cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; nervous system disorder; Alzheimer's disease;
 KW bacterial infection; viral infection; fungal infection; ocular disorder;
 KW corneal infection; wound healing; epithelial cell proliferation;
 KW skin aging; sunburn; organ transplantation; tissue regeneration;
 KW chemotaxis; food additive; food preservative; fat content; vitamin;
 KW mineral.

XX OS Homo sapiens.

XX XX US2004038277-A1.

XX PN 26-FEB-2004.

XX PD 18-JUL-2003; 2003US-00621401.

XX PF 07-NOV-1997; 97US-0064900P.

XX PR 07-NOV-1997; 97US-0064908P.

XX PR 07-NOV-1997; 97US-0064911P.

XX PR 07-NOV-1997; 97US-0064912P.

XX PR 07-NOV-1997; 97US-0064983P.

XX PR 07-NOV-1997; 97US-0064984P.

XX PR 07-NOV-1997; 97US-0064985P.

XX PR 07-NOV-1997; 97US-0064987P.

XX PR 07-NOV-1997; 97US-0064988P.

XX PR 17-NOV-1997; 97US-0066089P.

XX PR 17-NOV-1997; 97US-0066090P.

XX PR 17-NOV-1997; 97US-0066094P.

XX PR 17-NOV-1997; 97US-0066095P.

XX PR 17-NOV-1997; 97US-0066100P.

XX PR 04-NOV-1998; 98WO-US023435.

XX PR 05-MAY-1999; 99US-00305736.

XX PR 13-OCT-2000; 2000US-0239893P.

XX PR 12-MAR-2001; 2001US-00818683.

XX PR 12-OCT-2001; 2001US-00974879.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PA Rosen CA, Feng P, Ruben SM, Ebner R, Olsen H, Ni J, Wei Y;

XX PI Soppet DR, Moore PA, Kxaw H, Lafleur DM, Shi Y, Janat F;

XX PI Endress GA, Carter KC, Birse CE;

CC The invention relates to isolated nucleic acid molecules encoding human
 CC secreted proteins. The polynucleotides and polypeptides are used in
 CC diagnosing a pathological condition or susceptibility to a pathological
 CC condition. Antibodies to the polypeptides can also be used in alleviating
 CC symptoms associated with the disorders and in diagnostic immunoassays
 CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
 CC Disorders which are diagnosed or treated include autoimmune diseases
 CC (e.g. rheumatoid arthritis), hyperproliferative disorders (e.g. neoplasms
 CC of the breast or liver), cardiovascular disorders (e.g. cardiac arrest),
 CC cerebrovascular disorders (e.g. cerebral ischaemia), nervous system
 CC disorders (e.g. Alzheimer's disease), infections caused by bacteria,
 CC viruses and fungi and ocular disorders (e.g. corneal infection). The
 CC polypeptides can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues and in chemotaxis. The polypeptides can also be
 CC used as food additives or preservatives to increase or decrease storage
 CC capabilities and fat content, as vitamins, as minerals and as other
 CC nutritional components. This sequence represents a human secreted
 CC polynucleotide of the invention.

XX SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Query Match 42.7%; Score 1060.4; DB 12; Length 1543;
 Best Local Similarity 99.8%; Pred. No. 1.7e-263;
 Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 267 CTTTCGGCGTCATCCCGCTTCGAAAGCATCATCCAGAGTTCGACCGTGGTTAACTATAG 326
 DB 1 CTTTCGGCGTCATCCCGCTTCGAAAGCATCATCCAGAGTTCGACCGTGGTTAACTATAG 60
 QY 327 ATCAACACATCATCTTGCATCTCATGGGTTCTATCAATTTTAAATGCTTTGATGAAG 386
 DB 61 ATCAACACATCATCTTGCATCTCATGGGTTCTATCAATTTTAAATGCTTTGATGAAG 120
 QY 387 AGCATGTTATCCACTAGGAAGAAATAGTGGTGGTACTGTTTACCCAGAGTTCGATGAAC 446
 DB 121 AGCATGTTATCCACTAGGAAGAAATAGTGGTGGTACTGTTTACCCAGAGTTCGATGAAC 180
 QY 447 CGCTGGCCTTATTCATTTGGATTTTAAATACATTTGAACATACTGTTCACATAAGAGAGCT 506
 DB 181 CGCTGGCCTTATTCATTTGGATTTTAAATACATTTGAACATACTGTTCACATAAGAGAGCT 240
 QY 507 ATGTGTGTTCCCTTGACCAACCTTTTAGCGGCTTCATCTATATCTATCTCTCTTCCTGCTTAC 566
 DB 241 ATGTGTGTTCCCTTGACCAACCTTTTAGCGGCTTCATCTATATCTATCTCTCTTCCTGCTTAC 300
 QY 567 AAGAGAACTTTGGAACCAAGAGGAGGAGCTTTAGCTGCTTTTATGCTATTTGTATGTAAC 626
 DB 301 AAGAGAACTTTGGAACCAAGAGGAGGAGCTTTAGCTGCTTTTATGCTATTTGTATGTAAC 360
 QY 627 AGGCTACATATCTCGGTGAGTAGCTGGATCCCTTTGATATGAAGGAGGATGCTATTTTTCG 686
 DB 361 AGGCTACATATCTCGGTGAGTAGCTGGATCCCTTTGATATGAAGGAGGATGCTATTTTTCG 420
 QY 687 ACTTCAGTTTCACATCTATTTATGGGTAAATCTGTAAAAAATCGGTGAGTTCCTTTTGGAC 746
 DB 421 ACTTCAGTTTCACATCTATTTATGGGTAAATCTGTAAAAAATCGGTGAGTTCCTTTTGGAC 480
 QY 747 AATGTCGCTGCTTATCTATTTCTATATGCTCTCTGCTTTGGGTGGTATGATTTAT 806
 DB 481 AATGTCGCTGCTTATCTATTTCTATATGCTCTCTGCTTTGGGTGGTATGATTTAT 540
 QY 807 CATCAATCTTATTCACATGCTATTTGTGTGTTACTGATGAGAGATACAGCAAAAG 866
 DB 541 CATCAATCTTATTCACATGCTATTTGTGTGTTACTGATGAGAGATACAGCAAAAG 600
 QY 867 AGTCTACATAGCATATAGCATTTCTACATTTGGTGGTTAATATATCAATGAGAGATACC 926
 DB 601 AGTCTACATAGCATATAGCATTTCTACATTTGGTGGTTAATATATCAATGAGAGATACC 660
 QY 927 TTTTGTGGGATTCAGGCCAATCAGAAACAGTGAACACATGCGAGCTGCAGTGTCTTTTC 986

New nucleic acid molecule encoding one of 125 human secreted proteins for
 diagnosing, preventing, treating or ameliorating medical conditions and
 used as food additives or preservatives.

Claim 1; SEQ ID NO 133; 531pp; English.

ADG78372 standard; cDNA; 1543 BP.
 ADG78372;
 11-MAR-2004 (first entry)
 Human secreted protein cDNA #123.
 Human; secreted protein; gene; ss; immune disorder;
 haematopoietic disorder; hyperproliferative disorder; infectious disease;
 inflammatory disorder; HIV; anaemia; thrombocytopenia; bleeding; stroke;
 myocardial infarction; Addison's disease; rheumatoid arthritis;
 dermatitis; Grave's disease; multiple sclerosis; Glomerulonephritis;
 diabetes; graft-versus-host disease; inflammatory bowel disease; cancer;
 bacterial infection; viral infection; parasitic infection; osteoporosis;
 wound; atherosclerosis; Alzheimer's disease; Parkinson's disease; ulcer;
 food additive; food preservative.
 Homo sapiens.
 US2003211472-A1.
 13-NOV-2003.
 28-MAR-2001; 2001US-00818683.
 05-MAY-1999; 99US-00305736.
 (FENG/) FENG P.
 (RUBEN/) RUBEN S M.
 (ROSE/) ROSEN C A.
 (EBNER/) EBNER R.
 (OLSEN/) OLSEN H S.
 (NIJ/) NI J.
 (WEI/) WEI Y.
 (SOPP/) SOPPET D R.
 (MOORE/) MOORE P A.
 (KYAW/) KYAW H.
 (LAFLEUR/) LAFLEUR D W.
 (SHI/) SHI Y.
 (JANAT/) JANAT F.
 (ENDRESS/) ENDRESS G A.
 (CARTER/) CARTER K C.
 Feng P, Ruben SM, Rosen CA, Ebner R, Olsen HS, Ni J, Wei Y;
 Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
 Endress GA, Carter KC;
 WPI; 2004-051575/05.
 P-PSDB; ADG78500.
 New secreted nucleic acid for diagnosing, preventing or treating diseases
 associated with aberrant expression or activity of the polypeptide it
 encodes, e.g. cancer, human immunodeficiency virus, Parkinson's disease,
 or diabetes.
 Claim 1; SEQ ID NO 133; 377bp; English.
 The invention also relates to human secreted proteins and the nucleic
 acids encoding them. The proteins and nucleic acids are useful in
 diagnosing, preventing, prognosing or treating diseases or disorders
 associated with aberrant expression and/or activity of the secreted
 proteins, such as immune disorders, haematopoietic disorders,
 hyperproliferative disorders, infectious diseases or inflammatory
 disorders. In particular, the diseases or disorders are HIV, anaemia,
 thrombocytopenia, bleeding, stroke, myocardial infarction, Addison's
 disease, rheumatoid arthritis, dermatitis, Grave's disease, multiple
 sclerosis, glomerulonephritis, diabetes, graft-versus-host disease,
 inflammatory bowel disease, cancer, bacterial infections, viral
 infections, parasitic infections, osteoporosis, wounds, atherosclerosis,
 Alzheimer's disease, Parkinson's disease or ulcers. The sequences may
 also be used as food additives or preservatives, or for modulating
 mammalian mental or physical characteristics. The nucleic acids are also

CC used in chromosome mapping, in forensic biology or as molecular weight
 markers. This sequence represents cDNA encoding a human secreted protein
 of the invention.
 XX
 SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
 Query Match 42.7%; Score 1060.4; DB 12; Length 1543;
 Best Local Similarity 99.8%; Pred. No. 1.7e-263;
 Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 267 CTTCCGCGTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCCGCGGTTTAACTATAG 326
 DB 1 CTTCCGCGTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCCGCGGTTTAACTATAG 60
 QY 327 ATCAACACATCATCTTGATCTCATGGTCTCATGAATTTTAAATGGTTTGAAGAAG 386
 DB 61 ATCAACACATCATCTTGATCTCATGGTCTCATGAATTTTAAATGGTTTGAAGAAG 120
 QY 387 AGCATGGTATCCACTAGGAAGATAGTAGTGTACTGTTTACCAGGGTTGATGATAAC 446
 DB 121 AGCATGGTATCCACTAGGAAGATAGTAGTGTACTGTTTACCAGGGTTGATGATAAC 180
 QY 447 CGCTGCGCTTATTCATTGGATTTTAAATPACATTAACAATACTGTTTCAATTAAGACGCT 506
 DB 181 CGCTGCGCTTATTCATTGGATTTTAAATPACATTAACAATACTGTTTCAATTAAGACGCT 240
 QY 507 ATGTGTGTTCTTGCAACCACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 566
 DB 241 ATGTGTGTTCTTGCAACCACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 300
 QY 567 AAGAGAACTTTGGAAACCAAGAGCAGGACTTTTAGCTGCTGTTTATTCTATTGTATACC 626
 DB 301 AAGAGAACTTTGGAAACCAAGAGCAGGACTTTTAGCTGCTGTTTATTCTATTGTATACC 360
 QY 627 AGCTACATATCTCGGTGAGTAGTGTGATCTCTTGTATGATGAGGCACTTCTATTGTCG 686
 DB 361 AGCTACATATCTCGGTGAGTAGTGTGATCTCTTGTATGATGAGGCACTTCTATTGTCG 420
 QY 687 ACTTCAGTTCACATATCTATTTATGGTAAATCTGTAAAACTGGGTCAGTTTGTGGAC 746
 DB 421 ACTTCAGTTCACATATCTATTTATGGTAAATCTGTAAAACTGGGTCAGTTTGTGGAC 480
 QY 747 ATGTGCTGCTGCTTATCTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
 DB 481 ATGTGCTGCTGCTTATCTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 807 CATCAATCTTATTCACATGCACTGATTTGTTGTTTACTGATGAGAGATACAGCAAAAG 866
 DB 541 CATCAATCTTATTCACATGCACTGATTTGTTGTTTACTGATGAGAGATACAGCAAAAG 600
 QY 867 AGCTACATAGCATATAGCACTTCTACATGCTGGTTTAAATATATCAATGAGATACC 926
 DB 601 AGCTACATAGCATATAGCACTTCTACATGCTGGTTTAAATATATCAATGAGATACC 660
 QY 927 TTTTGTGGGATTCAGGCAATCAGAACAAAGTGAACACATGGCAGCTGCGGTTGCTTTCG 986
 DB 661 TTTTGTGGGATTCAGGCAATCAGAACAAAGTGAACACATGGCAGCTGCGGTTGCTTTCG 720
 QY 987 ATGTGCAAGCTTATGCTTCTTTCAGTATCTGAGAGACGATTAACAAAAACAAGATT 1046
 DB 721 ATGTGCAAGCTTATGCTTCTTTCAGTATCTGAGAGACGATTAACAAAAACAAGATT 780
 QY 1047 CCAGACCCCTTTCTTTTGGGTATCATCTAGCTGCGAGGCTGTGTCTTCTAGTGTAT 1106
 DB 781 CCAGACCCCTTTCTTTTGGGTATCATCTAGCTGCGAGGCTGTGTCTTCTAGTGTAT 840
 QY 1107 CTATTGTCTTATACAGGTTTACATTGCAACCAATGAGTGCGGAGGTTTATTCTATTGCGGA 1166
 DB 841 CTATTGTCTTATACAGGTTTACATTGCAACCAATGAGTGCGGAGGTTTATTCTATTGCGGA 900
 QY 1167 TACTGGTATGCAAAAAATACATTCATTCCATTTATTTGCAATGCTGCTGAGCATCAACCTAC 1226
 DB 901 TACTGGTATGCAAAAAATACATTCATTCCAAATTTATTTGCAATGCTGCTGAGCATCAACCTAC 960

XX 04-NOV-1998; 98WO-US023435.
XX
XX 07-NOV-1997; 97US-0064900P.
XX 07-NOV-1997; 97US-0064908P.
XX 07-NOV-1997; 97US-0064911P.
XX 07-NOV-1997; 97US-0064912P.
XX 07-NOV-1997; 97US-0064933P.
XX 07-NOV-1997; 97US-0064934P.
XX 07-NOV-1997; 97US-0064985P.
XX 07-NOV-1997; 97US-0064987P.
XX 07-NOV-1997; 97US-0064988P.
XX 17-NOV-1997; 97US-0066089P.
XX 17-NOV-1997; 97US-0066090P.
XX 17-NOV-1997; 97US-0066094P.
XX 17-NOV-1997; 97US-0066095P.
XX 17-NOV-1997; 97US-0066100P.
XX
XX (HUNA-) HUMAN GENOME SCI INC.
XX
XX Feng P, Carter KC, Endress GA, Rosen CA, Ruben SM, Janat F;
XX Ni J, Wei Y, Moore PA, Soppet DR, Kyaw H, Lafleur DW, Olsen HS;
XX Shi Y, Ebner R;
XX
XX WPI: 1999-337740/28.
XX P-PSDB: AAY27689, AAY27918, AAY27919, AAY27920, AAY27921, AAY27922,
XX AAY27923.
XX
XX New human secreted proteins and coding sequences useful for treating
XX disorders of the immune system and hyperproliferative disorders.
XX
XX Claim 1; Page 347; 507pp; English.
XX
XX This sequence represents a nucleic acid molecule which encodes a secreted
XX human protein. The gene number is given in the descriptor line. The gene
XX can be used to generate fusion proteins by linking to the gene to a human
XX immunoglobulin F portion (e.g. AAX84924) for increasing the stability of
XX the fused protein as compared to the human protein only. The invention
XX relates to 125 novel genes and their fragments (nucleic acid sequences:
XX AAX84933-485057; amino acid sequences AAY27567-127933) which are useful
XX for preventing, treating or ameliorating medical conditions e.g. by
XX protein or gene therapy. Also, pathological conditions can be diagnosed
XX by determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.
XX Specific uses are described for each of the 125 polynucleotides, based on
XX which tissues they are most highly expressed in (see AAX84933 for
XX described uses)
XX
XX Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
XX
XX Query Match 42.78; Score 1060.4; DB 2; Length 1543;
XX Best Local Similarity 99.84; Pred. No. 1.7e-263;
XX Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 267 CTTGCCGCTATCCGGTTCGAAAGCATCATCCACGAGTTCGACCGGTGGTTAACTATAG 326
XX 1 CTTGCCGCTATCCGGTTCGAAAGCATCATCCACGAGTTCGACCGGTGGTTAACTATAG 60
XX
XX 327 ATCAACACATCATCTTCGATCTCATGGTTCATGAATTTTAAATGGTTTGATGAAG 386
XX 61 ATCAACACATCATCTTCGATCTCATGGTTCATGAATTTTAAATGGTTTGATGAAG 120
XX
XX 387 AGCATGGTATCCACTAGGAAGAATAGTAGGTGGTACTGTTTACCAGGGTGGTATGAAC 446
XX 121 AGCATGGTATCCACTAGGAAGAATAGTAGGTGGTACTGTTTACCAGGGTGGTATGAAC 180
XX
XX 447 CGCTGGCCTTATTCATTTGGATTTTAAATACATTCACATACATGTTTCACATAAGACGCT 506
XX 181 CGCTGGCCTTATTCATTTGGATTTTAAATACATTCACATACATGTTTCACATAAGACGCT 240
XX
XX 507 ATGTGTGTTCTTCGACCAACTTTTAGCGCCCTTACATCTATATCTACTTTCTGCTTAC 566
XX 241 ATGTGTGTTCTTCGACCAACTTTTAGCGCCCTTACATCTATATCTACTTTCTGCTTAC 300
XX

OS Homo sapiens.
 PN WO200190304-A2.
 XX 29-NOV-2001.
 PD 18-MAY-2001; 2001WO-US016450.
 PF 19-MAY-2000; 2000US-0205515P.
 PR (HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX PI Birse CE, Rosen CA;
 XX DR WPI; 2002-122018/16.
 DR P-PSDB; ABB99441.
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders.
 PT
 PT
 PT
 XX
 PS Claim 4; SEQ ID NO 412; 2081pp + Sequence Listing; English.
 XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB9040-ABB9044) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune rheumatoid arthritis, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing and (e) neurological diseases such as cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
 Query Match 42.8%; Score 1061; DB 6; Length 1209;
 Best Local Similarity 99.8%; Pred. No. 1.1e-263;
 Matches 1061; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 267 CTTCCGCGTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCCCGTGGTTAACTATAG 326
 DB 1 CTTCCGCGTCATCCGCTTCGAAAGCATCATCCAGAGTTCGACCCCGTGGTTAACTATAG 60
 QY 327 ATCACACATCTTCGATCTCATGGTTCTATGAATTTTAAATGGTTTGTGTAAG 386
 DB 61 ATCACACATCTTCGATCTCATGGTTCTATGAATTTTAAATGGTTTGTGTAAG 120
 QY 387 AGCATGGTATCCATAGGAAGAATAGTAGTGTGTCTGTTTACCCAGGTTGATGATAAC 446
 DB 121 AGCATGGTATCCATAGGAAGAATAGTAGTGTGTCTGTTTACCCAGGTTGATGATAAC 180
 QY 447 CGTGGCTTTATTCATGGATTTTAAATACATGACATCACTGTTTCATAGAGAGCT 506
 DB 181 CGTGGCTTTATTCATGGATTTTAAATACATGACATCACTGTTTCATAGAGAGCT 240
 QY 507 ATGTGTGTCTTCGACCAACTTTAGCGGCTTATACATCTATATCTACTTCTGTTTAC 566
 DB 241 ATGTGTGTCTTCGACCAACTTTAGCGGCTTATACATCTATATCTACTTCTGTTTAC 300
 QY 567 AAGGAACTTTGGAAACCAAGAGCAGACTTTTAGCTGCTGTTTATTTGCTATTGTACC 626
 DB 301 AAGGAACTTTGGAAACCAAGAGCAGACTTTTAGCTGCTGTTTATTTGCTATTGTACC 360

QY 627 AGCTACATATCTCGGTGAGTCTGGATCCCTTTTGATTAATGAAGCATTTGCTATTTTGC 686
 DB 361 AGCTACATATCTCGGTGAGTCTGGATCCCTTTTGATTAATGAAGCATTTGCTATTTTGC 420
 QY 687 ACTTCAGTTTCACATCTATTTATGGTAAATCTCTTAAACTGGTTCAGTTTTTGGAC 746
 DB 421 ACTTCAGTTTCACATCTATTTATGGTAAATCTCTTAAACTGGTTCAGTTTTTGGAC 480
 QY 747 AATGTGCTGCTCTTATCTCTATTTCTATATGCTCTCTCTTTGGGGTGGTTATGATTAT 806
 DB 481 AATGTGCTGCTCTTATCTCTATTTCTATATGCTCTCTCTTTGGGGTGGTTATGATTAT 540
 QY 807 CATCAATCTTATTCACATGCTATTTGTTGTTTACTGATGAGAGATACAGCAAAAG 866
 DB 541 CATCAATCTTATTCACATGCTATTTGTTGTTTACTGATGAGAGATACAGCAAAAG 600
 QY 867 AGTCTACATAGCATATAGCACTTTCTACATTTGTGGTTTAAATATTTATCAATGCGATACC 926
 DB 601 AGTCTACATAGCATATAGCACTTTCTACATTTGTGGTTTAAATATTTATCAATGCGATACC 660
 QY 927 TTTTGTGGGATTCACAGCCCAATCAGAACAGAGTGAACACATGCGAGCTGCAGGTGCTTTGC 986
 DB 661 TTTTGTGGGATTCACAGCCCAATCAGAACAGAGTGAACACATGCGAGCTGCAGGTGCTTTGC 720
 QY 987 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAACAAAACAGAGTT 1046
 DB 721 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAACAAAACAGAGTT 780
 QY 1047 CCAGAGCCCTTTCTTTTGGGTGATCACTAGCTCAGAGTCTGTGTTCTTAGTGTCAT 1106
 DB 781 CCAGAGCCCTTTCTTTTGGGTGATCACTAGCTCAGAGTCTGTGTTCTTAGTGTCAT 840
 QY 1107 CTATTTGACTTAPACAGGTATCAATTCACCAATGAGTGGCAGGTTTATTCATTGTGGGA 1166
 DB 841 CTATTTGACTTAPACAGGTATCAATTCACCAATGAGTGGCAGGTTTATTCATTGTGGGA 900
 QY 1167 TACTGGGTATGAAAATAACATATCCAAATTTATTCATCAGTCTGAGAGTCAACCTAC 1226
 DB 901 TACTGGGTATGAAAATAACATATCCAAATTTATTCATCAGTCTGAGAGTCAACCTAC 960
 QY 1227 GACTTGGGTGCTTTCTTTTGTATCTACATATTTCTTGTATGATACCTTTCCAGCAGGCT 1286
 DB 961 GACTTGGGTGCTTTCTTTTGTATCTACATATTTCTTGTATGATACCTTTCCAGCAGGCT 1020
 QY 1287 TTGTTTCTGCATCAAAAATATCAAGCATGAAAGATTTTGT 1329
 DB 1021 TTGTTTCTGCATCAAAAATATCAAGCATGAAAGATTTTGT 1063

RESULT 7
 AAX85055
 ID AAX85055 standard; DNA; 1543 BP.
 XX
 AC AAX85055;
 XX
 DT 30-JUL-1999 (first entry)
 XX
 DE Human secreted protein gene No. 123.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 XX
 EN WO9924836-A1.
 XX
 PD 20-MAY-1999.

XX AAH33264;
AC
DT 03-SEP-2001 (first entry)
XX Human colon cancer antigen encoding cDNA SEQ ID NO:320.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
FN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI; 2001-235357/24.
DR P-PSDB; AAG73833.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 2436; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
Query Match 42.8%; Score 1061; DB 4; Length 1209;
Best Local Similarity 99.8%; Pred. No. 1.1e-263;
Matches 1061; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 267 CTTCCGCGTCATCCGCTCGAAGAGCATCATCCACGAGTTCGACCGGTGTTACTATAG 326
DB 1 CTTCCGCGTCATCCGCTCGAAGAGCATCATCCACGAGTTCGACCGGTGTTACTATAG 60
QY 327 ATCAACACATCATCTTGCATCTCATGGTTCATGAAATTTTAAATTTGGTTGATGAAAG 386
DB 61 ATCAACACATCATCTTGCATCTCATGGTTCATGAAATTTTAAATTTGGTTGATGAAAG 120
QY 387 AGCATGTATCCATAGGAGATAGTAGTGTCTACTGTTTACCCAGGTTGATGATTAAC 446
DB 121 AGCATGTATCCATAGGAGATAGTAGTGTCTACTGTTTACCCAGGTTGATGATTAAC 180
QY 447 CGTGGGCTTATTCATTTGAAATTTAAATATCATTTGAACATPACTGTTCATAAGAGCGT 506
DB 181 CGTGGGCTTATTCATTTGAAATTTAAATATCATTTGAACATPACTGTTCATAAGAGCGT 240

QY 507 ATGTGTGTTCTTTGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 566
DB 241 ATGTGTGTTCTTTGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 300
QY 567 AAGAGAACTTTGGAACCAAG 626
DB 301 AAGAGAACTTTGGAACCAAG 360
QY 627 AGGCTACATATCTCGGTGAGTAGTCTCTTGAATAAGAGGAGAGAGAGAGAGAGAGAG 686
DB 361 AGGCTACATATCTCGGTGAGTAGTCTCTTGAATAAGAGGAGAGAGAGAGAGAGAGAG 420
QY 687 ACTTCAGTTTCACATATCTCTTATGAGGTAATACTGTGTAATACTGGGTGAGTTTGGAC 746
DB 421 ACTTCAGTTTCACATATCTCTTATGAGGTAATACTGTGTAATACTGGGTGAGTTTGGAC 480
QY 747 AATGTGCTGCTCTTATCTCTTATGAGGTAATACTGTGTAATACTGGGTGAGTTTGGAC 806
DB 481 AATGTGCTGCTCTTATCTCTTATGAGGTAATACTGTGTAATACTGGGTGAGTTTGGAC 540
QY 807 CATCAATCTTATTCAGTGTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 866
DB 541 CATCAATCTTATTCAGTGTGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600
QY 867 AGTCTACATAGCATATAGCATCTTCTACATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 926
DB 601 AGTCTACATAGCATATAGCATCTTCTACATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 660
QY 927 TTTTGTGGGATTCAGGCAATCAGAAAGTGAACATGCGAGCTGCGAGTGTCTTTTGC 986
DB 661 TTTTGTGGGATTCAGGCAATCAGAAAGTGAACATGCGAGCTGCGAGTGTCTTTTGC 720
QY 987 ATTGTCTGCAAGCTTATGTTCTTGTGAGTATCTGAGAGCCGATTAACAAACAGAGTT 1046
DB 721 ATTGTCTGCAAGCTTATGTTCTTGTGAGTATCTGAGAGCCGATTAACAAACAGAGTT 780
QY 1047 CCAGACCCCTTTCTTTTGGGTTATCACTAGCTGCGAGTGTCTGTTCTTCTTGTGTTGTTG 1106
DB 781 CCAGACCCCTTTCTTTTGGGTTATCACTAGCTGCGAGTGTCTGTTCTTCTTGTGTTGTTG 840
QY 1107 CTATTGTGCTATACAGGTTTACATTTGCACTGAGTGGAGGTTTATTCTATTGTTGTTGTTG 1166
DB 841 CTATTGTGCTATACAGGTTTACATTTGCACTGAGTGGAGGTTTATTCTATTGTTGTTGTTG 900
QY 1167 TACTGGGTATGCAAAATACATCTTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1226
DB 901 TACTGGGTATGCAAAATACATCTTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 1227 GACTTGGGTGCTTCTTCTTGTGATCTACATATTTCTTGTATGTTGTTGTTGTTGTTGTTG 1286
DB 961 GACTTGGGTGCTTCTTCTTGTGATCTACATATTTCTTGTATGTTGTTGTTGTTGTTGTTG 1020
QY 1287 TTTGTTCTGCTCAAAATATCAACGATGAAAGAGTATTTGTT 1329
DB 1021 TTTGTTCTGCTCAAAATATCAACGATGAAAGAGTATTTGTT 1063

RESULT 6
ABL89850
ID ABL89850 standard; cDNA; 1209 BP.
XX
AC ABL89850;
XX
DT 24-MAY-2002 (first entry)
XX
XX Human polynucleotide SEQ ID NO 412.
XX
XX Cytostatic; immunosuppressive; noctropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antineoplastic; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.

XX DT 20-MAY-2004 (first entry)
 XX DE Full length human cDNA clone SeqID 2694.
 XX KW human; medicine; signal transduction; glycoprotein; transcription;
 XX KW oligo-capping method; ss; gene.
 XX OS Homo sapiens.
 XX PN EP1396543-A2.
 XX PD 10-MAR-2004.
 XX PF 07-JUL-2000; 2003EP-00025638.
 XX PR 08-JUL-1999; 99JP-00194486.
 XX PR 11-JAN-2000; 2000JP-00118774.
 XX PR 02-MAY-2000; 2000JP-001183865.
 XX PR 07-JUL-2000; 2000EP-00114089.
 XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX DR WPI; 2004-204755/20.
 XX DR P-PSDB; ADL30662.
 XX PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
 XX PT length human cDNAs.
 XX FS Example 1; SEQ ID NO 2694; 1340pp; English.
 XX CC This invention relates to a novel primers useful for synthesizing full
 XX CC length cDNA molecules that encode human proteins. Specifically, it refers
 XX CC to secretory or membrane proteins that are potential therapeutic agents/
 XX CC target molecules in the field of medicine, and in particular genes
 XX CC encoding proteins that are associated with signal transduction,
 XX CC glycoproteins and transcription. The present invention describes a method
 XX CC for efficiently cloning a full length human cDNA from both the 5' and 3',
 XX CC ends using the oligo-capping method. This polynucleotide sequence is a
 XX CC full length human cDNA clone of the invention.
 XX SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;
 Query Match 53.6%; Score 1328.8; DB 12; Length 1664;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGCGAGCCCTCGGCGCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGG 60
 DB 28 ATGGCGAGCCCTCGGCGCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGG 87
 QY 61 AGTGGCTCATAGCCCTTGGGAAACAGCCGCAAGCCGCAAGCCGCGCGCGCGG 120
 DB 88 AGTGGCTCATAGCCCTTGGGAAACAGCCGCAAGCCGCAAGCCGCGCGCGCGG 147
 QY 121 GCGCACAGCGG 180
 DB 148 GCGCACAGCGG 207
 QY 181 GGGCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 240
 DB 208 GGGCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 267
 QY 241 TGGCTTCCGCGCTTACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 300
 DB 268 TGGCTTCCGCGCTTACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 327
 QY 301 GAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTGCATCTCANGGTTCTAT 360
 DB 328 GAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTGCATCTCANGGTTCTAT 387

QY 361 GAATTTTAAATGGTTTTCATGAAAGAGCATGGTATCCACTAGGAAGAAATAGTAGTGGT 420
 DB 388 GAATTTTAAATGGTTTTCATGAAAGAGCATGGTATCCACTAGGAAGAAATAGTAGTGGT 447
 QY 421 ACTGTTTACCCAGGTTGATGATAACCGCTGGCCTTATTCATTGGATTTTAAATACATTG 480
 DB 448 ACTGTTTACCCAGGTTGATGATAACCGCTGGCCTTATTCATTGGATTTTAAATACATTG 507
 QY 481 AACATAACTGTTTACATAGAGAGCATGATGTGTGTTCCCTTGACCAACATTTTAGCGGCTT 540
 DB 508 AACATAACTGTTTACATAGAGAGCATGATGTGTGTTCCCTTGACCAACATTTTAGCGGCTT 567
 QY 541 ACATCTATATCTACTTTCTGCTTCAAGAGAGACTTTGGAAACCAAGAGGAGGACTTTTA 600
 DB 568 ACATCTATATCTACTTTCTGCTTCAAGAGAGACTTTGGAAACCAAGAGGAGGACTTTTA 627
 QY 601 GCTGCTGTTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTGAGTGTGATCTCTTT 660
 DB 628 GCTGCTGTTGTTTATTTGCTATTGTACAGGCTACATATCTCGGTGAGTGTGATCTCTTT 687
 QY 661 GATAATGAGGCAATGCTATTTTTCGACTTCAGTTTTCATATCTATTTATGCTGTAATCT 720
 DB 688 GATAATGAGGCAATGCTATTTTTCGACTTCAGTTTTCATATCTATTTATGCTGTAATCT 747
 QY 721 GTAAAAAAGTGGTCAAGTTTGTGCAATGCTGCTCTTATCTCTATTTCTATATGCTC 780
 DB 748 GTAAAAAAGTGGTCAAGTTTGTGCAATGCTGCTCTTATCTCTATTTCTATATGCTC 807
 QY 781 TCTGCTTGGGTTGTTATGTTATTTATCATCATCTTATTCACATGCTGCTGATTTGTGTTG 840
 DB 808 TCTGCTTGGGTTGTTATGTTATTTATCATCAATCTTATTCACATGCTGCTGATTTGTGTTG 867
 QY 841 TTACTGATCAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTCTTACATTGTG 900
 DB 868 TTACTGATCAGAGATACAGCAAAAGAGTCTACATAGCATATAGCACCTTCTTACATTGTG 927
 QY 901 GGTTTTAAATTAATGAGATACCTTTTGTGGGATTCAGGCAATCAGAACAAAGTGA 960
 DB 928 GGTTTTAAATTAATGAGATACCTTTTGTGGGATTCAGGCAATCAGAACAAAGTGA 987
 QY 961 CACATGGCAGCTGCGAGGTGCTTTGCTGCTGCAAGCTTATGCTTCTTGCAGTATCTG 1020
 DB 988 CACATGGCAGCTGCGAGGTGCTTTGCTGCTGCAAGCTTATGCTTCTTGCAGTATCTG 1047
 QY 1021 AGAGACCGATTAACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTTGATCATAGCT 1080
 DB 1048 AGAGACCGATTAACAAAACAGAGTTCAGACCCCTTTTCTTTTGGGTTGATCATAGCT 1107
 QY 1081 GCAGGTGCTGTGTTCTTGTCTAGTCTCATCTATTGTACTTATACAGGTTACATTGCACCATGG 1140
 DB 1108 GCAGGTGCTGTGTTCTTGTCTAGTCTCATCTATTGTACTTATACAGGTTACATTGCACCATGG 1167
 QY 1141 AGTGGCAGGTTTATTCATTGTGGGATAGTGGGATGCAAAAATACATTCCTCAATTAAT 1200
 DB 1168 AGTGGCAGGTTTATTCATTGTGGGATAGTGGGATGCAAAAATACATTCCTCAATTAAT 1227
 QY 1201 GCATCAGTGTCTGAGCATCAACTGACTTGGGTTGCTTCTTCTTGTGATCTACATATT 1260
 DB 1228 GCATCAGTGTCTGAGCATCAACTGACTTGGGTTGCTTCTTCTTGTGATCTACATATT 1287
 QY 1261 CTTGTATGATCTTCCAGCAGGCTTTGGTTCTGCTCATCAAAAATATCAACGATGAAGA 1320
 DB 1288 CTTGTATGATCTTCCAGCAGGCTTTGGTTCTGCTCATCAAAAATATCAACGATGAAGA 1347
 QY 1321 GTATTTGTTGCT 1332
 DB 1348 GTATTTGTTGCT 1359

RESULT 5
 AAH33264
 ID AAH33264 standard; cDNA; 1209 BP.

DE Human full-length cDNA, SEQ ID NO: 2694.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 KW Homo sapiens.
 OS EP1130094-A2.
 XX EP1130094-A2.
 XX 05-SEP-2001.
 PD 07-JUL-2000; 2000EP-00114089.
 PF 08-JUL-1999; 99JP-00194486.
 XX 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI WPI; 2001-524255/58.
 DR P-PSDB; AAM93252.
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PT Claim 8; SEQ ID NO 2694; 1380pp + Sequence Listing; English.
 PS The invention relates to primers for synthesizing full length cDNA
 XX clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5' and 3' ends of the cDNA molecules have
 CC been determined. Primers for synthesizing the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesized by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO
 XX
 SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;
 Query Match 53.6%; Score 1328.8; DB 4; Length 1664;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGCAAGTCGTCCTCAACTCGPCCCGGTGG 60
 28 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGTCGTCCTCAACTCGPCCCGGTGG 87
 61 AGTGGCTCATAGCCCTGGGAAACAGCGGCGACGGCCACACCGGCGCGGCGCGGCGG 120
 88 AGTGGCTCATAGCCCTGGGAAACAGCGGCGACGGCCACACCGGCGCGGCGCGGCGG 147
 121 GCGCACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 148 GCGCACAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207
 181 GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 208 GGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 267
 241 TGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 268 TGGCTGTGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 327
 301 GAGTTCGACCGCGGCTTAACTATAGATCAACACATCATCTCTGATCTCATCGGCTTCAT 360
 328 GAGTTCGACCGCGGCTTAACTATAGATCAACACATCATCTCTGATCTCATCGGCTTCAT 387
 361 GAATTTTAAATGGTTTGGATGAAGAGCATGATGATCCATAGGAAGATAGTAGGTGGT 420

| | | |
|------|---|------|
| 388 | GAATTTTAAATGGTTTGGATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGGT | 447 |
| 421 | ACTGTTTACCAGGGTTGATGATAAACCCTGCGCTTATTCATTCGATTTTAAATACATTC | 480 |
| 448 | ACTGTTTACCAGGGTTGATGATAAACCCTGCGCTTATTCATTCGATTTTAAATACATTC | 507 |
| 481 | AACATAAATGTTTCACATAAGAGAGCATGATGATGCTTCCTGACCAACTTTTAGCGGCTT | 540 |
| 508 | AACATAAATGTTTCACATAAGAGAGCATGATGATGCTTCCTGACCAACTTTTAGCGGCTT | 567 |
| 541 | ACATCTATATCTACTTCTCTGCTTACAGAGAACTTTTGGACCAAGAGAGAGAGCTTTTA | 600 |
| 568 | ACATCTATATCTACTTCTCTGCTTACAGAGAACTTTTGGACCAAGAGAGAGAGCTTTTA | 627 |
| 601 | GCTGCTTGTGTTTATTTGCTTATTTGCTTACAGGCTTACATATCTCGGTCAGTAGTCCCTT | 660 |
| 628 | GCTGCTTGTGTTTATTTGCTTATTTGCTTACAGGCTTACATATCTCGGTCAGTAGTCCCTT | 687 |
| 661 | GATATGAGGAGCATTTGCTTATTTGCTTACAGGCTTACATATCTCGGTCAGTAGTCCCTT | 720 |
| 688 | GATATGAGGAGCATTTGCTTATTTGCTTACAGGCTTACATATCTCGGTCAGTAGTCCCTT | 747 |
| 721 | GTAAATACTGGTCAAGTTTGTGCAATGTGCTGCTTATCTATTTCTATATGCTG | 780 |
| 748 | GTAAATACTGGTCAAGTTTGTGCAATGTGCTGCTTATCTATTTCTATATGCTG | 807 |
| 781 | TCGCTTGGGGTGTGATGATTTATCATCAATCTTATCCACGTCATGATTTGCTG | 840 |
| 808 | TCGCTTGGGGTGTGATGATTTATCAATCTTATCCACGTCATGATTTGCTG | 867 |
| 841 | TTACTGATGACAGATACAGCAAAAGAGCTCTACATAGCATATAGCATTTCTCATGTTG | 900 |
| 868 | TTACTGATGACAGATACAGCAAAAGAGCTCTACATAGCATATAGCATTTCTCATGTTG | 927 |
| 901 | GGTTTAAATATATCAATGACAGATACCTTTGCGGATTCAGGCAATCAGACAGAGTAA | 960 |
| 928 | GGTTTAAATATATCAATGACAGATACCTTTGCGGATTCAGGCAATCAGACAGAGTAA | 987 |
| 961 | CACATGGCAGCTGCGAGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAGTATCTG | 1020 |
| 988 | CACATGGCAGCTGCGAGGTGCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAGTATCTG | 1047 |
| 1021 | AGAGACCGATTAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCT | 1080 |
| 1048 | AGAGACCGATTAACAAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCT | 1107 |
| 1081 | GCAGGTGCTGCTGCTTCTTCTAGTGTCTATCTATTTGACTTATACAGGTTACATTGACCATGG | 1140 |
| 1108 | GCAGGTGCTGCTGCTTCTTCTAGTGTCTATCTATTTGACTTATACAGGTTACATTGACCATGG | 1167 |
| 1141 | AGTGGCAGGTTTATTCATTTGTTGGATCTGGGTATGCAAAATATACATCTCCATTTAT | 1200 |
| 1168 | AGTGGCAGGTTTATTCATTTGTTGGATCTGGGTATGCAAAATATACATCTCCATTTAT | 1227 |
| 1201 | GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTTGATCTACATATT | 1260 |
| 1228 | GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTTGATCTACATATT | 1287 |
| 1261 | CTTGTATGATCTTCCAGAGGCGCTTGGTCTGCAATCAAAATATCAACGATGAAGA | 1320 |
| 1288 | CTTGTATGATCTTCCAGAGGCGCTTGGTCTGCAATCAAAATATCAACGATGAAGA | 1347 |
| 1321 | GTATTTTGTGCT | 1332 |
| 1348 | GTATTTTGTGCT | 1359 |

RESULT 4
 ADL30661
 ID ADL30661 standard; cDNA; 1664 BP.
 XX
 AC ADL30661;

QY 481 AACATAACTGTTCATAGAGAGCGTATGTGTTCCTTGCACCAACTTTTAGCGCCCTT 540
DB 548 AACATAACAGTTCACATAAGAGATGTGTGTATTCCTTGCACCAACTTTTAGCGCCCTT 607
QY 541 ACATCTATCTACTTTCTCTTACAGAGAACTTTGGAACCAAGGACGAGCTTTTA 600
DB 608 ACATCCATACTACGTTCTCTTACAGAGAACTTTGGAACCAAGGACGAGCTTTTA 667
QY 601 GCTGCTTTGTTTATGTATGTACAGGCTACATATCTCGGTCAGTGTGATGCTGAT 660
DB 668 GCTGCTGCTTCATTGTCTATCGTACAGGCTACATATCTCGGTCAGTGTGATGCTG 727
QY 661 GATATGAGGCTATGCTATTTTGGCACTTCAGTTCACATATTTTATGAGTAACT 720
DB 728 GATATGAGGCTATGCTATTTTGGCACTTCAGTTCACATATTTTATGAGTAACT 787
QY 721 GTAAAACTGGGTGAGTATTTTGGCACTTCAGTTCACATATTTTATGAGTAACT 780
DB 788 GTGAAGACCGGCTGCTGTTCTGACAAATGTGCTGCTGTGCTATTTTACATGCTC 847
QY 781 TCTGCTTGGGTGCTGCTATTTTATCATCAATCTTATTTCCAGTGTATTTTGTG 840
DB 848 TCTGCTTGGGTGCTGCTATTTTATCATCAATCTTATTTCCAGTGTATTTTGTG 907
QY 841 TTACTGATGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTTACATGCT 900
DB 908 CTGCTGATGAGATACAGCAAAAGAGTCTACATAGCATATAGCACTTTTACATGCT 967
QY 901 GGTATATATATCAATGCGAGTACCTTTTGGGATTCAGGCTTCAGGCTTCAGGCT 960
DB 968 GGTATATATATCAATGCGAGTACCTTTTGGGATTCAGGCTTCAGGCTTCAGGCT 1027
QY 961 CACATGCGAGTGTGCTTGTGCTATGCTGCAAGCTTATGCTTTTGTGAGTATCTG 1020
DB 1028 CACATGCGAGTGTGCTTGTGCTATGCTGCAAGCTTATGCTTTTGTGAGTATCTG 1087
QY 1021 AGAGACGATTAACAAACAGAGTTCAGGCTTCAGGCTTCAGGCTTCAGGCT 1080
DB 1088 AGAGACGATTAACAAACAGAGTTCAGGCTTCAGGCTTCAGGCTTCAGGCT 1147
QY 1081 GCAGGTGTGTGTTTCTTGTGCTATCTATGACATACAGGCTTATGCAAGCTT 1140
DB 1148 GCAGGTGTGTGTTTCTTGTGCTATCTATGACATACAGGCTTATGCAAGCTT 1207
QY 1141 AGTGGAGTATTAATCAATGCGAGTTCAGGCTTCAGGCTTCAGGCTTCAGGCT 1200
DB 1208 AGTGGAGTATTAATCAATGCGAGTTCAGGCTTCAGGCTTCAGGCTTCAGGCT 1267
QY 1201 GCATCAGTGTGAGCATCAACCTACGATGCTGCTTCTTCTTGTGATCTACATAT 1260
DB 1268 GCATCAGTGTGAGCATCAACCTACGATGCTGCTTCTTCTTGTGATCTACATAT 1327
QY 1261 CTGATGATGCTTCCAGAGGCTTGTGCTGATCAATCAAAATACAGATGAAGA 1320
DB 1328 CTGATGATGCTTCCAGAGGCTTGTGCTGATCAATCAAAATACAGATGAAGA 1387
QY 1321 GTATTTGCTGCTATATGCAATCAGTGTGCTGCTTCTTCTTGTGATGCTGCTG 1380
DB 1388 GTATTTGCTGCTATATGCAATCAGTGTGCTGCTTCTTCTTGTGATGCTGCTG 1447
QY 1381 ATGTTGATGCTTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1448 ATGTTGATGCTTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507
QY 1441 GAGCATATTTGGGGATGATGATGAAGGGAATTCACCTGTGAGGACAGAGTAT 1500
DB 1508 GAGCATATTTGGGGATGATGATGAAGGGAATTCACCTGTGAGGACAGAGTAT 1567
QY 1501 GAGGATGACAAAGAAACCAAGGAAATTTGATGATGAGGAGTAAAGTGAAGAACT 1560
DB 1568 GAGGATGACAAAGAAACCAAGGAAATTTGATGATGAGGAGTAAAGTGAAGAACT 1627

QY 1561 GCAACTGAACAGGAAACCTGAGAGGATTTAGCCCTTAATATATAAAGCAATGTCTACC 1620
DB 1628 GTGACAGAGCAAGAGAAACCTGAAGAGGCTTTGGGCCCAACATCAAAAGCAATGTGTACC 1687
QY 1621 ATCTTGTATCTGTATGCTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1688 ATCTTGTATCTGTATGCTATTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1747
QY 1681 GCCTACTGTAGTCCAAAGTGTAGTCTGCTGCTCATACAATCATATGATGCGACAGAAATATC 1740
DB 1748 GCCTACTGTAGTCCAAAGTGTAGTCTGCTGCTCATACAATCATATGATGCGACAGAAATAT 1807
QY 1741 TTAGATGATTTAGAGAACTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCAAGCA 1800
DB 1808 TTAGATGATTTAGAGAACTTACTTTTGGCTTAAGCAAAATACAGATGAACATGCAAGCA 1867
QY 1801 GTATGCTTGTGGGATTTATGCTATCAGATGATGCTGAATGGCTTAATPAGAACTACGTTG 1860
DB 1868 GTATGCTTGTGGGATTTATGCTATCAGATGATGCTGAATGGCTTAATPAGAACTACGTTG 1927
QY 1861 GTGATTAATAACACCTGGAATACAGCCACATAGCACTGCTGGGAAAGCTATGTCTTCT 1920
DB 1928 GTGATTAATAACACCTGGAATACAGCCACATAGCACTGCTGGGAAAGCTATGTCTTCT 1987
QY 1921 AATGAACACAGGCTTATTAATCATATGAGGCTCTAGATGATGATTTGTTGTTTAT 1980
DB 1988 AATGAACACAGGCTTATTAATCATATGAGGCTCTAGATGATGATTTGTTGTTTAT 2047
QY 1981 TTTGGAGGCTTATTTGGCTTATTTCTGCTGATGATATCAAAATTTCTCTGATGCTGTTAGG 2040
DB 2048 TTTGGAGGCTTATTTGGCTTATTTCTGCTGATGATATCAAAATTTCTCTGATGCTGTTAGG 2107
QY 2041 ATAGCTGAAGGAGAAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCCAAGGGA 2100
DB 2108 ATAGCTGAAGGAGAAACATCCCAAGACATTCGGGAAAGTACTATTTTACCCCAAGGGA 2167
QY 2101 GAATTCCTGTAGACAAAGCAGGATCCCTACTTTTGTGATGCTGCTTATGATTAATG 2160
DB 2168 GAATTCCTGTAGACAAAGCAGGATCCCTACTTTTGTGATGCTGCTTATGATTAATG 2227
QY 2161 TCATCTACAGATTTTGAGAAATGACGCTGGAATTTTCTGACACCCCAAGGTTTACCGGA 2220
DB 2228 TCATCTACAGATTTTGAGAAATGACGCTGGAATTTTCTGACACCCCAAGGTTTACCGGA 2287
QY 2221 ACAGTAAAGCTGAGATTTGGAATGAGGACATTAATTCAAACATTTGGAAGGACCTTT 2280
DB 2288 ACAGTAAAGCTGAGATTTGGAATGAGGACATTAATTCAAACATTTGGAAGGACCTTT 2347
QY 2281 ACATCAGAACACTGGCTTGTAGGATATATAAAGTAAAGGACCTGATAACAGGAGAGA 2340
DB 2348 ACATCAGAACACTGGCTTGTAGGATATATAAAGTAAAGGACCTGATAACAGGAGAGA 2407
QY 2341 TTAGATCAAAACCTGAGTCAACACATTTTCCAAACAGAGATTTTGTCAAGAGAG 2400
DB 2408 CTAGGTCAAAACCTGAGTCAACACATTTTCCAAACAGAGATTTTGTCAAGAGAG 2467
QY 2401 ACTACCAAAAGGAGGCTGCTGCTACATTAATAAAGTAAAGGCTGTTTAAAGAAAGGCAAGAAA 2460
DB 2468 ACTACCAAAAGGAGGCTGCTGCTACATTAATAAAGTAAAGGCTGTTTAAAGAAAGGCAAGAAA 2527
QY 2461 ATATCTAAGNAGACTGTTTAA 2481
DB 2528 ACCTCTAAGNAGACTGTTTAA 2548

RESULT 3

AAK94164

ID AAK94164 standard; cDNA; 1664 BP.

XX

AC AAK94164;

XX

DT 06-NOV-2001 (first entry)

XX

Db 1801 GTAATGCTCTTGGTGGGATATGCTATCATAGATAGCTGGAATGGCTAATAGAACTACGTTG 1850
Qy 1861 GTGGATATTAACACCTGGATATACAGCCATAGCACTGCTGGGAAAGCTATGCTTCT 1920
Db 1861 GTGGATATTAACACCTGGATATACAGCCATAGCACTGCTGGGAAAGCTATGCTTCT 1920
Qy 1921 AATGAACACAGCCTATAAAATCATGAGGACTCTAGATGTAGATATGTTTGGTTAT 1980
Db 1921 AATGAACACAGCCTATAAAATCATGAGGACTCTAGATGTAGATATGTTTGGTTAT 1980
Qy 1981 TTTGGAGGGTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGATGTTAGG 2040
Db 1981 TTTGGAGGGTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGATGTTAGG 2040
Qy 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGA 2100
Db 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGA 2100
Qy 2101 GAATTCGGTGTAGCAAGCAGGATCCCTACTTGTGATTTGCTTGAATGCTTATGTAATG 2160
Db 2101 GAATTCGGTGTAGCAAGCAGGATCCCTACTTGTGATTTGCTTGAATGCTTATGTAATG 2160
Qy 2161 TCATCTACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGGTTTTCACCGA 2220
Db 2161 TCATCTACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGGTTTTCACCGA 2220
Qy 2221 ACAGTAACTGCTGATGGAATGGAATAGGACATTAATTCAAACATTTGGAAGAGCCTTT 2280
Db 2221 ACAGTAACTGCTGATGGAATGGAATAGGACATTAATTCAAACATTTGGAAGAGCCTTT 2280
Qy 2281 ACATCAGAACACTGGCTTTGTAGGATATATAAGTAAAGACACCTGTATACAGGGAGACA 2340
Db 2281 ACATCAGAACACTGGCTTTGTAGGATATATAAGTAAAGACACCTGTATACAGGGAGACA 2340
Qy 2341 TTAGATCACAACTCGAGTCACACATTTTCCGAAACAGAGATTTTGTCAAGAG 2400
Db 2341 TTAGATCACAACTCGAGTCACACATTTTCCGAAACAGAGATTTTGTCAAGAG 2400
Qy 2401 ACTACCAAAAGGAGCGTGCTACATTAATAAATAAGCTGTTTAAAGAAAGCGAAGAA 2460
Db 2401 ACTACCAAAAGGAGCGTGCTACATTAATAAATAAGCTGTTTAAAGAAAGCGAAGAA 2460
Qy 2461 ATATCTAAGAGACTGTTTAA 2481
Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 2

ADD94785

ID ADD94785 standard; cDNA; 2710 BP.

AC ADD94785;

XX ADD94785;

DT 29-JAN-2004 (first entry)

XX 29-JAN-2004 (first entry)

DE Mouse SIMP cDNA sequence.

XX Mouse SIMP cDNA sequence.

KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
KW murine; gene; ss.

XX Mus musculus.

OS Mus musculus.

XX WO2003054008-A2.

PN 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 18-DEC-2002; 2002WO-CA001967.

PF 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.
PR (COMP-) COMPATIGENE INC.
PA Perreault C, McBride K;
XX WPI; 2003-559122/52.
DR P-PSDB; ADD94786.
XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
PT or breast cancer, or for suppressing an immune response in an autoimmune
PT disease.
XX Disclosure; SEQ ID NO 3; 66pp; English.
XX This invention relates to a novel isolated or purified human protein,
CC termed source of immunodominant major histocompatibility complex (MHC) -
CC associated peptide (SIMP), which is expressed ubiquitously in human
CC cells, where the protein has the potential of generating several protein
CC fragments binding with high affinity to a human leukocyte antigen (HLA)
CC molecule. The invention may allow development of therapeutics with
CC cytostatic or immunosuppressive activity or provide sequences useful for
CC antisense therapy or gene therapy. The source of immunodominant MHC-
CC associated peptide (SIMP) nucleic acids, proteins and fragments are
CC useful for diagnosing and treating cancers, for example lung cancer,
CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
CC proteins are also useful for modulating an immune response. Decreasing
CC lymphoid cell proliferation is useful for suppressing an immune response
CC responsible for an autoimmune disease or a transplant rejection. The
CC present sequence is that of the mouse SIMP cDNA which is related to the
CC invention. Note: This sequence does not appear in the specification but
CC was obtained by the indexer from GenBank.

XX Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

Query Match 83.8%; Score 2078.6; DB 10; Length 2710;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 234; Indels 9; Gaps 1;

Qy 1 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGCAGCAGTCTCCCTCACTCTCCCTCGG 60
Db 77 ATGGCGGAGCCCTCGGCGCGGAGAGCAAGCAGCAGTCTCCCTCACTCTCCCTCGG 136
Qy 61 AGTGGGCTCATGCGCCCTCGGAAACAGCGGACAGCCACCGGCGCGGCGCGGCGG 120
Db 137 AGCGGCTCATGCGCTCTGGGAAACAGCGGACAGCCACCGGCGCGGCGGCGGAGC 196
Qy 121 GCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
Db 197 GCGTCCAGGCGG-----GCGGCGGCGGAAAGCGGCGGCGGCGGCGGCGGCGG 247
Qy 181 GGGCTGTGCGAGCGCGGCTGGGTGGGCGAGTCTCTCTCTTCACTCTCTCTCTG 240
Db 248 GGGCTGTGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 307
Qy 241 TGGCTTCCGCGGCTCAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 308 TGGCTTCCGCGGCTCAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 367
Qy 301 GAGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 368 GAGTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427
Qy 361 GAATTTTAAATGGTTGATGAAGAGCAGCATGATCCACTAGGAGAGTAGTGGT 420
Db 428 GAGTTTCTAAATGGTTGATGAAGAGCAGCATGATCCACTAGGAGAGTAGTGGTGGC 487
Qy 421 ACTGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCTATTGATTTAAATACATTG 480
Db 488 ACCGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCTATTGATTTAAATACATTG 547

termed source of immunodominant major histocompatibility complex (MHC) - associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcoma, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the human SIMP cDNA which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

XX SQ Sequence 2481 BP: 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Query Match 100.0%; Score 2481; DB 10; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGAGCGCTCGCGCGGAGCAGCAGCAAGTCGCTCAACTCGTCCCGTGG 60
 Db 1 ATGCGGAGCGCTCGCGCGGAGCAGCAGCAAGTCGCTCAACTCGTCCCGTGG 60

QY 61 AGTGGCTCATGGCCCTGGGAAACAGCGCGCAGCCAGCGGCGCGGCGCCAGTGC 120
 Db 61 AGTGGCTCATGGCCCTGGGAAACAGCGCGCAGCCAGCGGCGCGGCGCCAGTGC 120

QY 121 GCGCACAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 180
 Db 121 GCGCACAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 180

QY 181 GGGCTGTCGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 240
 Db 181 GGGCTGTCGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 240

QY 241 TGGCTTCGCGGCTCAGCTCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCT 300
 Db 241 TGGCTTCGCGGCTCAGCTCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCT 300

QY 301 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTGCACTCATGGTTCAT 360
 Db 301 GAGTTCGACCGCGTGTAACTATAGATCAACACATCATCTGCACTCATGGTTCAT 360

QY 361 GAATTTTAAATTTGATGAAGAGCATGTATCCATAGGAGATAGTAGTGGT 420
 Db 361 GAATTTTAAATTTGATGAAGAGCATGTATCCACTAGGAGATAGTAGTGGT 420

QY 421 ACTGTTTACCGAGGTTGATGAACCGCGTGGCGCTTATTCATTTGGAATTTAAATG 480
 Db 421 ACTGTTTACCGAGGTTGATGAACCGCGTGGCGCTTATTCATTTGGAATTTAAATG 480

QY 481 AACATACTGTCACATAGAGCATGTATGTTGCTTCCGACCACTTTAGCGGCTT 540
 Db 481 AACATACTGTCACATAGAGCATGTATGTTGCTTCCGACCACTTTAGCGGCTT 540

QY 541 ACATCTATATCTACTTTCCCTGCTTACAAGAGAACTTTTGGAAACCAAGGACGAGCTTTA 600
 Db 541 ACATCTATATCTACTTTCCCTGCTTACAAGAGAACTTTTGGAAACCAAGGACGAGCTTTA 600

QY 601 GCTGCTGTTTATGCTATGTACAGGCTACATATCGCTCAGTAGCTGGATCCCTT 660
 Db 601 GCTGCTGTTTATGCTATGTACAGGCTACATATCGCTCAGTAGCTGGATCCCTT 660

QY 661 GATAATGAAGGCATTTGCTATTTTGGCACTTCAGTTCACATATTTATGGTAAATCT 720
 Db 661 GATAATGAAGGCATTTGCTATTTTGGCACTTCAGTTCACATATTTATGGTAAATCT 720

QY 721 GTAAABAACTGGTCAAGTCTTTTGGCAAAAGTGGCTGCTTATCTATATGTC 780

Db 721 GTAAABAACTGGTCAAGTCTTTTGGCAAAAGTGGCTGCTTATCTATATGTC 780
 QY 781 TCTGCTTGGGCTGGTGTATGTTATTTATCATCAATCTTATCCACGCAATGTTGTTG 840
 Db 781 TCTGCTTGGGCTGGTGTATGTTATTTATCATCAATCTTATCCACGCAATGTTGTTG 840

QY 841 TTACTGATCAGAGATACAGCAAGAGCTCTACATAGCATATAGCATTCTTACATGTC 900
 Db 841 TTACTGATCAGAGATACAGCAAGAGCTCTACATAGCATATAGCATTCTTACATGTC 900

QY 901 GGTATAATTTATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGA 960
 Db 901 GGTATAATTTATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTGA 960

QY 961 CACATCGAGCTGCGAGTCTTTCATTTGCTGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020
 Db 961 CACATCGAGCTGCGAGTCTTTCATTTGCTGCTGCAAGCTTATGCTTCTTCAGTATCTG 1020

QY 1021 AGAGACCGATTAAACAAACAGAGTTCAGACCCCTTTTCTTTGGGTTATCACTAGCT 1080
 Db 1021 AGAGACCGATTAAACAAACAGAGTTCAGACCCCTTTTCTTTGGGTTATCACTAGCT 1080

QY 1081 GCAGGTGCTGTGTCCTTAGTCTCATCTATTTGACTTATACAGGTTACATTCGACCATG 1140
 Db 1081 GCAGGTGCTGTGTCCTTAGTCTCATCTATTTGACTTATACAGGTTACATTCGACCATG 1140

QY 1141 AGTGGCAGGTTTATTTATTTGTTGGATATCGGTTATGCAAAATACACATTTCAATTA 1200
 Db 1141 AGTGGCAGGTTTATTTATTTGTTGGATATCGGTTATGCAAAATACACATTTCAATTA 1200

QY 1201 GCATCAGTCTGAGCATCAACTAGCTTGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1260
 Db 1201 GCATCAGTCTGAGCATCAACTAGCTTGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1260

QY 1261 CTTGTATGTAACCTTCCAGCAGCGCTTTGGTGTCTGCAATCAAAATATCAACGATGAAGA 1320
 Db 1261 CTTGTATGTAACCTTCCAGCAGCGCTTTGGTGTCTGCAATCAAAATATCAACGATGAAGA 1320

QY 1321 GTATTTTGTGCTCTATATGCAATCAGTCTGCTACTTTGCTGAGTATGCTGCGAGCTG 1380
 Db 1321 GTATTTTGTGCTCTATATGCAATCAGTCTGCTACTTTGCTGAGTATGCTGCGAGCTG 1380

QY 1381 ATGTTGACTTTGACTCCAGTCTGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTT 1440
 Db 1381 ATGTTGACTTTGACTCCAGTCTGTGTATGCTGTCTGCAATTTGCTTTTCAAAATGTTT 1440

QY 1441 GAGCCTATTTGGGGATGACATGAAGGGAATCCACCTGTGGAGGACAGCAGTGTAT 1500
 Db 1441 GAGCCTATTTGGGGATGACATGAAGGGAATCCACCTGTGGAGGACAGCAGTGTAT 1500

QY 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTGAAGAAACAT 1560
 Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGCGAGGTAAAGTGAAGAAACAT 1560

QY 1561 GCACTGACAGGAAACCAAGGAAATTTGATGATAGGCGAGGTAAAGTGAAGAAACAT 1620
 Db 1561 GCACTGACAGGAAACCAAGGAAATTTGATGATAGGCGAGGTAAAGTGAAGAAACAT 1620

QY 1621 ATGTTGATGCTGATGCTATTTGATGATGTTGCTGCTCACTGTACTGCTGCTGCTGCT 1680
 Db 1621 ATGTTGATGCTGATGCTATTTGATGATGTTGCTGCTCACTGTACTGCTGCTGCTGCT 1680

QY 1681 GCCTACTCTAGTCCAGTGTAGTCTGCTCATCAATCATGATGGCAGGCAATATC 1740
 Db 1681 GCCTACTCTAGTCCAGTGTAGTCTGCTCATCAATCATGATGGCAGGCAATATC 1740

QY 1741 TTAGATGATTTAGAGAACTTTATTTTGGTGAAGCAAAATACAGATGAACATGACGA 1800
 Db 1741 TTAGATGATTTAGAGAACTTTATTTTGGTGAAGCAAAATACAGATGAACATGACGA 1800

QY 1801 GTAACTCTGTTGGTGAATATGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTG 1860

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 20:50:43 ; Search time 1138.18 Seconds
(without alignments)
11442.663 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcggagccctcgccccc.....tatctaaagagactgtttaa 2481

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseqn23Sep04:*
2: Geneseqn1380s:*
3: Geneseqn1390s:*
4: Geneseqn2000s:*
5: Geneseqn2001as:*
6: Geneseqn2001bs:*
7: Geneseqn2002as:*
8: Geneseqn2002bs:*
9: Geneseqn2003as:*
10: Geneseqn2003bs:*
11: Geneseqn2003cs:*
12: Geneseqn2003ds:*
13: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 2481 | 100.0 | 2481 | 10 | ADD94783 Human SIM |
| 2 | 2078.6 | 83.8 | 2710 | 10 | ADD94783 Mouse SIM |
| 3 | 1328.8 | 53.6 | 1664 | 4 | AAK94164 Human ful |
| 4 | 1328.8 | 53.6 | 1664 | 12 | ADL30661 Full leng |
| 5 | 1061 | 42.8 | 1209 | 4 | AAH33284 Human col |
| 6 | 1061 | 42.8 | 1209 | 6 | ABL98850 Human pol |
| 7 | 1060.4 | 42.7 | 1543 | 2 | AAH85055 Human sec |
| 8 | 1060.4 | 42.7 | 1543 | 8 | ACD18981 Novel hum |
| 9 | 1060.4 | 42.7 | 1543 | 12 | ADG78372 Human sec |
| 10 | 1060.4 | 42.7 | 1543 | 12 | ADN60663 Human sec |
| 11 | 972 | 39.2 | 2546 | 2 | AAV44866 Clone CT5 |
| 12 | 972 | 39.2 | 2546 | 5 | AAH98453 Human cdn |
| 13 | 955.8 | 38.5 | 2417 | 10 | ADD94789 Drosophil |
| 14 | 955.8 | 38.5 | 2699 | 4 | ABL13247 Drosophil |
| 15 | 954 | 38.5 | 2510 | 5 | AAH33772 Human cdn |
| 16 | 892.4 | 36.0 | 2537 | 4 | AAH08289 Human sec |
| 17 | 892.4 | 36.0 | 2537 | 4 | AAH08315 Human sec |
| 18 | 892 | 36.0 | 2650 | 6 | ABO54750 Human ova |
| 19 | 867.2 | 35.0 | 1114 | 4 | AAH59794 Human pro |
| 20 | 861.4 | 34.7 | 2284 | 4 | AAH18021 Human cdn |
| 21 | 781.5 | 30.7 | 4922 | 4 | ABL13246 Drosophil |

| | | | | | |
|----|-------|------|------|----|--------------------|
| 22 | 712.8 | 28.7 | 787 | 4 | AAH07526 Human cdn |
| 23 | 689.2 | 27.8 | 2236 | 6 | ABZ32036 Candida a |
| 24 | 671 | 27.0 | 764 | 5 | AAH93968 Primer sp |
| 25 | 670.6 | 27.0 | 3094 | 10 | ADD94791 Mouse itm |
| 26 | 655 | 26.4 | 2855 | 4 | ABL02795 Drosophil |
| 27 | 646.6 | 26.1 | 2472 | 10 | ADD94793 Human itm |
| 28 | 646.6 | 26.1 | 2472 | 10 | ADH28844 Human chr |
| 29 | 629.2 | 25.4 | 2760 | 5 | ABV24502 Human pro |
| 30 | 609.4 | 24.6 | 2466 | 10 | ADB89900 C. neofo |
| 31 | 606.2 | 24.4 | 2232 | 8 | ABT20789 Aspergill |
| 32 | 537.8 | 21.7 | 2733 | 10 | ADD94787 Yeast STT |
| 33 | 525.6 | 21.2 | 2000 | 10 | ACC61293 Gere sequ |
| 34 | 525.6 | 21.2 | 2000 | 10 | ADK63659 Disease t |
| 35 | 515 | 20.8 | 1848 | 8 | ABT18969 Aspergill |
| 36 | 507 | 20.4 | 507 | 3 | AAC01967 Human sec |
| 37 | 500 | 20.2 | 500 | 6 | ABV88332 Human col |
| 38 | 488 | 18.9 | 503 | 4 | AAK91644 Human cdn |
| 39 | 468 | 18.9 | 503 | 4 | AAK93577 Human cdn |
| 40 | 468 | 18.9 | 503 | 12 | ADL30004 3' end of |
| 41 | 468 | 18.9 | 503 | 12 | ADL28071 5' end of |
| 42 | 458.2 | 18.5 | 558 | 5 | AAH93612 Umbilical |
| 43 | 456.4 | 18.4 | 1969 | 8 | AAH18375 Aspergill |
| 44 | 456.4 | 18.4 | 2603 | 8 | ABT20191 Aspergill |
| 45 | 456.4 | 18.4 | 3969 | 8 | ABT17781 Aspergill |

ALIGNMENTS

RESULT 1

ADD94783
ID ADD94783 standard; cdna; 2481 BP.

AC ADD94783;

DT 23-JAN-2004 (first entry)

DE Human SIMP cdna sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
XX major histocompatibility complex; human leukocyte antigen; HLA;
XX cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
XX lung cancer; intestine cancer; sarcoma; prostate cancer;
XX testicular cancer; breast cancer; melanoma; pancreatic cancer;
XX haematological cancer; immune response; lymphoid cell proliferation;
XX autoimmune disease; transplant rejection; SIMP-derived peptide; human;
XX gene; ss.
XX Homo sapiens.
XX WO2003054008-A2.
XX 03-JUL-2003.
XX 18-DEC-2002; 2002WO-CA001967.
XX 20-DEC-2001; 2001US-00028384.
XX (COMP-) COMPATIGENE INC.
XX Perreault C, McBride K;
XX WPI: 2003-559122/52.
XX P-PSDB; ADD94784.

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Claim 6; SEQ ID NO 1; 66pp; English.

This invention relates to a novel isolated or purified human protein.


```

DEFINITION 602618240F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4732256 5',
            mRNA sequence.
ACCESSION  BG620025
VERSION     BG620025.1 GI:13671396
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 776)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM1591 row: o column: 09
            High quality sequence stop: 727.
            Location/Qualifiers
            1..776
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:4732256"
            /lab_host="NIH_MGC_79"
            /note="Organ: Placenta; Vector: pDNR-LIB (Clontech);
            Site 1: Sfii (ggcgctctggcc); Site 2: Sfii
            (ggcattatggcc); 5' and 3' adaptors were used in cloning
            as follows: 5' adaptor sequence: 5'-CACGCCCATTTAGGCC-3'
            and 3' adaptor sequence:
            5'-ATTCTAGGCGGAGGCGGCGGCATG-dt(30)BN-3' (where B = A,
            C, or G and N = A, C, G, or T). Average insert size 1.3
            kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
            by PCR. This library was enriched for full-length clones
            and was constructed by Clontech Laboratories (Palo Alto,
            CA). Note: this is a NIH_MGC Library."

ORIGIN
Query Match      20.8%; Score 517; DB 4; Length 776;
Best Local Similarity 99.7%; Pred. No. 1e-266;
Matches 617; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1863 GGATAATAACACCTGGGAATACAGCCACATAGACACTGGTGGGAAAAGCTATGCTTTCAA 1922
DB 1 GGATAATAACACCTGGGAATACAGCCACATAGACACTGGTGGGAAAAGCTATGCTTTCAA 60

QY 1923 TGAACAGCAGCCTATAAATCATGAGACTCTAGATGATAGATTGTTTGGTTATTTT 1982
DB 61 TGAACAGCAGCCTATAAATCATGAGACTCTAGATGATAGATTGTTTGGTTATTTT 120

QY 1983 TGGAGGGGTTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGGATGTTAGGAT 2042
DB 121 TGGAGGGGTTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGGATGTTAGGAT 180

QY 2043 AGCTGAGGAGAAATCCCAAGACATTCGGGAAGTGACTATTTACCCACAGGAGA 2102
DB 181 AGCTGAGGAGAAATCCCAAGACATTCGGGAAGTGACTATTTACCCACAGGAGA 240

QY 2103 ATTCGCTGAGACAAAGCAGGATCCCTACTTGTGTTGAATGCTTATGATAAAATGTC 2162
DB 241 ATTCGCTGAGACAAAGCAGGATCCCTACTTGTGTTGAATGCTTATGATAAAATGTC 300

QY 2163 ATACTACAGATTGGAGAAATGAGCTGGATTTTCGTACACCCCGAGTTTGACCGAAC 2222
DB 301 ATACTACAGATTGGAGAAATGAGCTGGATTTTCGTACACCCCGAGTTTGACCGAAC 360

```

```

QY 2223 ACCTAATGCTGAGATTGGAATTAAGGACATTAAATTCAAACATTGGAAGAAGCCTTTAC 2282
DB 361 ACCTAATGCTGAGATTGGAATTAAGGACATTAAATTCAAACATTGGAAGAAGCCTTTAC 420

QY 2283 ATCAGAACACTGGCTGTTTGGATATATAAAGTAAAGCACTGATACAGGGAGACATT 2342
DB 421 ATCAGAACACTGGCTGTTTGGATATATAAAGTAAAGCACTGATACAGGGAGACATT 480

QY 2343 AGATCAACAACTCGAGTCACCAACATTTCCTCCAAAACAGAAAGTATTTGTCAAAAGAGAC 2402
DB 481 AGATCAACAACTCGAGTCACCAACATTTCCTCCAAAACAGAAAGTATTTGTCAAAAGAGAC 540

QY 2403 TACCAAAAGGAGCGTGGCTACATTAAATAAAGCTGCTGTTTAAAGAAAGCAAGAAAT 2462
DB 541 TACCAAAAGGAGCGTGGCTACATTAAATAAAGCTGCTGTTTAAAGAAAGCAAGAAAT 600

QY 2463 ATCTAAGAAGACTGTTTAA 2481
DB 601 ATCTAAGAAGACTGTTTAA 619

RESULT 15
CD103829 793 bp mRNA linear EST 15-MAY-2003
LOCUS     AGENCOURT_14007965 NIH_MGC_186 Homo sapiens cDNA clone
DEFINITION IMAGE:30372391 5', mRNA sequence.
ACCESSION  CD103829
VERSION     CD103829
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovics
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM164 row: i column: 08
            High quality sequence stop: 374.
            Location/Qualifiers
            1..793
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:30372391"
            /lab_host="DH10B (TI phage-resistant)"
            /note="Organ: Pooled-Skin; Vector: pDNR-LIB; Site 1: Sfii
            (ggcattatggcc); Site 2: Sfii (ggcgctctggcc); Library is
            oligo-dt primed and directionally cloned. cDNA was
            prepared from a pooled samples of tissues from Skin,
            meninges, duramater, pia matter and choroid plexus.
            and 3' adaptors were used in cloning as follows: 5'
            adaptor sequence: 5'-CACGCCCATTTAGGCC-3' and 3' adaptor
            sequence: 5'-ATTCTAGGCGGAGGCGGCGGCATG-dt(30)BN-3'
            (where B = A, C, or G and N = A, C, G, or T). Average
            insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies
            contained inserts by PCR. This library was enriched for
            full-length clones and was constructed by Clontech
            Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
            Library"

FEATURES
source

```

ORIGIN

Query Match

20.5%; Score 508; DB 6; Length 793;

QY 1794 TGCACGAGTAATGCTTTGGGATATGCTATCAGATAGCTGGAAATGGCTAATAGAAC 1853
 Db 361 TGCACGAGTAATGCTTTGGGATATGCTATCAGATAGCTGGAAATGGCTAATAGAAC 420
 QY 1854 TACCTTGGTGAATATACACCTTGGATACACCCACATAGACACTGGTGGGAAAGCTAT 1913
 Db 421 TACCTTGGTGAATATACACCTTGGATACACCCACATAGACACTGGTGGGAAAGCTAT 480
 QY 1914 CTCCTTCAATGAACACACGACCTATAAAATCATGAGAGCTCTAGATGTAGATATGTTTT 1973
 Db 481 CTCCTTCAATGAACACGACGCTATAAAATCATGAGAGCTCTAGATGTAGATATGTTTT 540
 QY 1974 GGTATATTT 1982
 Db 541 GGTATATTT 549

RESULT 12
 BG434322 627 bp mRNA linear EST 14-MAR-2001
 LOCUS 602506307F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4603584 5',
 DEFINITION mRNA sequence.

ACCESSION BG434322.1 GI:13340828

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L10M1337 row: n column: 01

High quality sequence stop: 625.

Location/Qualifiers

1. 627

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="IMAGE:4603584"

/lab_hosts="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 79"

/note="Organ: placenta; Vector: pDNR-LIB (Clontech);

Site 1: SfiI (ggcgctcgcc); Site 2: SfiI

(ggcgataggcc); 5' and 3' adaptors were used in cloning

as follows: 5' adaptor sequence: 5'-CACGCCATTAGGCC-3'

and 3' adaptor sequence:

5'-ATTCTAGGCCGCGGCGGCACATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.3

kb (range 0.5-4.0 kb). 15/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match

Best Local Similarity 22.0%; Score 546; DB 4; Length 627;

Matches 546; Conservativeness 0; Mismatches 0; Indels 0; Gaps 0;

QY 621 TGTACCGGCTACATATCTGGTCAGTAGCTGGATCCTTTGATAATGAAGGATGCTAT 680

Db 81 TGTACCGGCTACATATCTGGTCAGTAGCTGGATCCTTTGATAATGAAGGATGCTAT 140

QY 681 TTTTGCACTTCAGTTCCATCTATTTATGGGTAATCTGTAAAACTGGGTGAGTTT 740
 Db 141 TTTTGCACTTCAGTTCCATCTATTTATGGGTAATCTGTAAAACTGGGTGAGTTT 200
 QY 741 TTGGCAATGTCTGCTGCTTATCCCTATTTCTATTTCTATTTCTGCTGCTGGGTGATGT 800
 Db 201 TTGGCAATGTCTGCTGCTTATCCCTATTTCTATTTCTATTTCTGCTGCTGGGTGATGT 260
 QY 801 ATTATCATCAATCTTTATTTCCACTGCATGTATTTCTGTTTACTGATGCAGATACAG 860
 Db 261 ATTATCATCAATCTTTATTTCCACTGCATGTATTTCTGTTTACTGATGCAGATACAG 320
 QY 861 CAAAAGAGCTCATAGCATATAGACATTTCTACATTTCTGCTGCTTAAATATTAATCAATGCA 920
 Db 321 CAAAAGAGCTCATAGCATATAGACATTTCTACATTTCTGCTGCTTAAATATTAATCAATGCA 380
 QY 921 GATACCTTTTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGAGTGCAGTGT 980
 Db 381 GATACCTTTTGGGATTCAGCCCAATCAGAACAGTGAACACATGCGAGTGCAGTGT 440
 QY 981 CTTTGCATTTGCTGCAAGCTTATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1040
 Db 441 CTTTGCATTTGCTGCAAGCTTATGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
 QY 1041 AGAGTTCCAGACCCCTTTCTTTTGGGTATCATCTAGCTGCGAGTGCCTTCTTCTTCTTCT 1100
 Db 501 AGAGTTCCAGACCCCTTTCTTTTGGGTATCATCTAGCTGCGAGTGCCTTCTTCTTCTTCTTCT 560
 QY 1101 TGTCACTATTGACTTATACAGGTTTACATTTCCACCATGGAGTGCAGGTTTATTCATT 1160
 Db 561 TGTCACTATTGACTTATACAGGTTTACATTTCCACCATGGAGTGCAGGTTTATTCATT 620
 QY 1161 GTGGGA 1166
 Db 621 GTGGGA 626

RESULT 12

LOCUS CN263069

DEFINITION 17000600180485 GRN_PRENU Homo sapiens cDNA 5', mRNA sequence.

ACCESSION CN263069

VERSION CN263069.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 577)

Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Murage, J., Fisk, G.J.,

Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,

Lebkowski, J. and Stanton, L.W.

Transcriptome characterization elucidates signaling networks that

control human ES cell growth and differentiation

Nat. Biotechnol. 22 (6), 707-716 (2004)

Contact: Brandenberger R

Regenerative Medicine

Gen Corporation

230 Constitution Drive, Menlo Park, CA 94025, USA

Tel: 650 473 8658

Fax: 650 473 7760

Email: rbrandenberger@gen.com

Insert length: 577, Std Error: 0.00.

Location/Qualifiers

1. 577

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="embryonic stem cell, retinoic acid and

mitogen-treated hES cell line H7"

/clone_lib="GRN_PRENU"

/note="oligo dt primed, full-length enriched cDNA library

Plate: 5 row: E column: 12
High quality sequence stop: 562.

FEATURES

Source
1. 562
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L118NU354-5-E12"
/sex="M"
/tissue_type="Liver"
/cell_type="Polygonal"
/lab_hosts="Topi0r"
/clone_lib="L118NU354"
/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including
EcoRI site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 50nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Topi0r by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

ORIGIN

Query Match 22.7%; Score 562; DB 6; Length 562;
Best Local Similarity 100.0%; Pred. No. 5.1e-291;
Matches 562; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1643 TGATGTTTGTCTGCTCACTGACCTGGGTCAACAGCAATGCTTACTCTAGTCCAAAGTGTAG 1702
DB 1 TGATGTTTGTCTGCTCACTGACCTGGGTCAACAGCAATGCTTACTCTAGTCCAAAGTGTAG 60
QY 1703 TCTTGGCTCATACAAATCATGATGGCCACAGGAATATCTAGATGATTTTACAGAGCTT 1762
DB 61 TCTTGGCTCATACAAATCATGATGGCCACAGGAATATCTAGATGATTTTACAGAGCTT 120
QY 1763 ACTTTTGGCTAAGCAAAATACAGATGAACATGACGAGTAATGCTTGGTGGGATTAAG 1822
DB 121 ACTTTTGGCTAAGCAAAATACAGATGAACATGACGAGTAATGCTTGGTGGGATTAAG 180
QY 1823 GCTATCAGATAGCTGGAAATGCTTAATAGAACTACGTTGGTGGATTAATACACTGGAATA 1882
DB 181 GCTATCAGATAGCTGGAAATGCTTAATAGAACTACGTTGGTGGATTAATACACTGGAATA 240
QY 1883 ACAGCCATAGCACTGGTGGGAAAAGCTATGCTTCTTAATGAACACAGCAGCTTAATAA 1942
DB 241 ACAGCCATAGCACTGGTGGGAAAAGCTATGCTTCTTAATGAACACAGCAGCTTAATAA 300
QY 1943 TCATGAGGACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGTTATTGGCTATT 2002
DB 301 TCATGAGGACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGTTATTGGCTATT 360
QY 2003 CTGTGTATGATATCAACAAATTTCTCTGGATGTTTAGGATAGCTGAAGGAGAACTCCCA 2062
DB 361 CTGTGTATGATATCAACAAATTTCTCTGGATGTTTAGGATAGCTGAAGGAGAACTCCCA 420
QY 2063 AAGACATTTGGGAAAGTGACTATTTTACCCACAGGGAGAAATTCCTGTGTAGCAAGAG 2122
DB 421 AAGACATTTGGGAAAGTGACTATTTTACCCACAGGGAGAAATTCCTGTGTAGCAAGAG 480
QY 2123 GATCCCTCATTTTGGATTTGCTTATGTATTAATATGATACATACAGATTTGGAGAAA 2182
DB 481 GATCCCTCATTTTGGATTTGCTTATGTATTAATATGATACATACAGATTTGGAGAAA 540
QY 2183 TGCAGCTGGATTTTTCGTACACC 2204

Db 541 TGCAGCTGGATTTTTCGTACACC 562

RESULT 10

EG698273 735 bp mRNA linear EST 07-MAY-2001
LOCUS 602650188F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4803332 5',
DEFINITION mRNA sequence.
ACCESSION EG698273 GI:13965393
VERSION EG698273.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 735)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10698 row: p column: 21
High quality sequence start: 2
High quality sequence stop: 734.

FEATURES

Location/Qualifiers
1. 735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4803332"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP Skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 22.1%; Score 549; DB 4; Length 735;
Best Local Similarity 100.0%; Pred. No. 5.4e-284;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1434 TGTTTTGGACCTATTTGGGGATGACATGAAGGAAATCCACTGTGGAGCAG 1493
DB 1 TGTTTTGGACCTATTTGGGGATGACATGAAGGAAATCCACTGTGGAGCAG 60
QY 1494 CAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTATGAAGCAGTAAAGTGAG 1553
DB 61 CAGTGTAGGATGACAAAGAAACCAAGGAAATTTGTATGAAGCAGTAAAGTGAG 120
QY 1554 GACATGCAACTGACAGGAAACCTGAAGAGGATTAGCCCTAATAAAGCAT 1613
DB 121 GAACATGCAACTGACAGGAAACCTGAAGAGGATTAGCCCTAATAAAGCAT 180
QY 1614 TGTCAACATGTTGCTGATGCTATTGATGATGTTTGTCTGCACTGTACCTGGGTAC 1673
DB 181 TGTCAACATGTTGCTGATGCTATTGATGATGTTTGTCTGCACTGTACCTGGGTAC 240
QY 1674 AGCATGCTACTCTAGTCCCAAGTGTAGTCTGSCCTCATACATCATGATGACACAG 1733
DB 241 AGCAATGCTACTCTAGTCCCAAGTGTAGTCTGSCCTCATACATCATGATGACACAG 300
QY 1734 GAATATCTTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACA 1793
DB 301 GAATATCTTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACA 360

QY 1678 AATGCTACTCTAGTCCAGTGTAGTCTCTGGCTCATACATCATGATGCGCACCAGGAAT 1737
 Db 241 AATGCTACTCTAGTCCAGTGTAGTCTCTGGCTCATACATCATGATGCGCACCAGGAAT 300
 QY 1738 ATCTTAGATGATTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAATGCA 1797
 Db 301 ATCTTAGATGATTTAGAGAGCTTACTTTTGGCTTAAGGCAAAATACAGATGAATGCA 360
 QY 1798 CGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGC 1857
 Db 361 CGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGC 420
 QY 1858 TTGGTGATGATTAACACCTGGATTAACAGCAGCATAGCACTGGTGGGAAAGCTATGCT 1917
 Db 421 TTGGTGATGATTAACACCTGGATTAACAGCAGCATAGCACTGGTGGGAAAGCTATGCT 480
 QY 1918 TCTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTT 1777
 Db 481 TCTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTT 540
 QY 1978 ATTTTGGAGGGTATTGGCTATTGCTATTCGGTGATGATATCAAC 2019
 Db 541 ATTTTGGAGGGTATTGGCTATTGCTATTCGGTGATGATATCAAC 582

RESULT 8
 BQ877881
 LOCUS
 DEFINITION BQ877881 889 bp mRNA linear EST 16-AUG-2002
 AGENCOURT 7981106 Lupski dorsal root ganglion Homo sapiens cDNA
 clone IMAGE:6185364 5', mRNA sequence.

ACCESSION BQ877881
 VERSION BQ877881.1 GI:22269889
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 889)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: rcapbs@email.nih.gov
 Tissue Procurement: Dr. James R. Lupski
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM13576 row: a column: 13
 High quality sequence stop: 673.

FEATURES
 Location/Qualifiers
 1..889
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6185364"
 /sex="male"
 /tissue_type="dorsal root ganglia"
 /dev_stage="adult, 36 yr"
 /lab_host="DH10B"
 /clone_lib="Lupski dorsal root ganglion"
 /note="vector: pCMV-SPORT6 (Life Technologies); Site_1:
 Note; Site 2: SalI; cDNA made by oligo-dt priming.
 Directionally cloned using the following adaptors:
 5'-TCGACCCACGGCCG-3' and
 5'-GATAGTCTAGTCCGCGCGCCGCT(15)-3'. Size selected >
 1 kb for average insert length 1.7 kb. This is a primary
 library, non-amplified. Library constructed by Life
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
 College of Medicine) and is available through Life

Technologies."

Query Match 23.5%; Score 582; DB 5; Length 889;
 Best Local Similarity 100.0%; Pred. No. 8.5e-302;
 Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1900 GTGGGAAAAGCTATGCTTTCTAATGAAAAGCAGGCTTATAAATCATGAGACTCTAGAT 1959
 Db 1 GTGGGAAAAGCTATGCTTTCTAATGAAAAGCAGGCTTATAAATCATGAGACTCTAGAT 60
 QY 1960 GTAGATTATGTTTGGTTATTTTGGAGGGGTATTGGCTATTCTGGTGATGATATCAAC 2019
 Db 61 GTAGATTATGTTTGGTTATTTTGGAGGGGTATTGGCTATTCTGGTGATGATATCAAC 120
 QY 2020 AATTTCTCTGATGTTAGGATAGCTGAAGAGAAACATCCCAAAGACATTCGGGAAAGT 2079
 Db 121 AATTTCTCTGATGTTAGGATAGCTGAAGAGAAACATCCCAAAGACATTCGGGAAAGT 180
 QY 2080 GACTATTTTACCCACAGGGGAGAAATCCGTGTAGACAAAGCAGGATCCCTACTTTGTTG 2139
 Db 181 GACTATTTTACCCACAGGGGAGAAATCCGTGTAGACAAAGCAGGATCCCTACTTTGTTG 240
 QY 2140 AATGCTTATGATATAAATGTCATACAGATTTGGAGAAATGCGAGCTGGATTTTCGT 2199
 Db 241 AATGCTTATGATATAAATGTCATACAGATTTGGAGAAATGCGAGCTGGATTTTCGT 300
 QY 2200 ACACCCCGAGGTTTGGACCGAACAGCTAATGCTGAGATTGGAAATTAAGGACATTAATTC 2259
 Db 301 ACACCCCGAGGTTTGGACCGAACAGCTAATGCTGAGATTGGAAATTAAGGACATTAATTC 360
 QY 2260 AAACATTTGGAAGAGCCCTTTACATCAGAACTGCTGGCTTTAGGATATATAAGATAAAA 2319
 Db 361 AAACATTTGGAAGAGCCCTTTACATCAGAACTGCTGGCTTTAGGATATATAAGATAAAA 420
 QY 2320 GCACCTGTAAACAGGGAGACATTTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAA 2379
 Db 421 GCACCTGTAAACAGGGAGACATTTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAA 480
 QY 2380 CAGAGATTTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAAAAAATAGCTG 2439
 Db 481 CAGAGATTTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAAAAAATAGCTG 540
 QY 2440 GTTTTAAAGAAAGGCAAGAAATATCTAAGAGAGCTGTTTAA 2481
 Db 541 GTTTTAAAGAAAGGCAAGAAATATCTAAGAGAGCTGTTTAA 582

RESULT 9
 BQ128931
 LOCUS
 DEFINITION K-EST0178489 L11SNU354 Homo sapiens cDNA clone L11SNU354-5-E12 5',
 mRNA sequence.
 ACCESSION BQ128931
 VERSION BQ128931.1 GI:28092254
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 562)
 Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
 Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
 Kim, Y.S.
 21C Frontier Korean EST Project 2001
 Unpublished (2002)
 Contact: Kim YS
 Genom Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr

```

RESULT 6
CD358477      903 bp      mRNA      linear      EST 29-MAY-2003
LOCUS      AGNCOURT_14252099 NIH_MGC_180 Homo sapiens cDNA clone
DEFINITION      IMAGE:30388860 5', mRNA sequence.
ACCESSION      CD358477
VERSION      CD358477.1 GI:31129912
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 903)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Blodg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapos-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM461 row: 9 column: 13
High quality sequence stop: 641.
Location/Qualifiers
1..903
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30388860"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
Note: Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

FEATURES
source
1..903
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30388860"
/lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
/clone_lib="NIH_MGC_180"
Note: Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
Site 2: EcoRV (destroyed); Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.68 kb. Library was
constructed by (Invitrogen). Note: this is a NIH_MGC
Library."

ORIGIN
Query Match      23.8%; Score 590; DB 6; Length 903;
Best Local Similarity 100.0%; Pred. No. 4.1e-306;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1872 CACCTGAATACAGCCACATAGACACTGGTGGAAAGCTATGCTTCTTAATGAACACG 1931
DB 60 CACCTGAATACAGCCACATAGACACTGGTGGAAAGCTATGCTTCTTAATGAACACG 119
QY 1932 AGCCTATAAATCATAGGACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGGGT 1991
DB 120 AGCCTATAAATCATAGGACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGGGT 179
QY 1992 TATTGGCTATTCGGTATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGG 2051
DB 180 TATTGGCTATTCGGTATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGG 239
QY 2052 AGAACATCCCAAGAACATTCGGGAAAGTGACTATTTTACCCACAGGAGGAATTCGGTGT 2111
DB 240 AGAACATCCCAAGAACATTCGGGAAAGTGACTATTTTACCCACAGGAGGAATTCGGTGT 299
QY 2112 AGACAAAGCAGGATCCCTACTTTTGTGAATTCGCTTATGATATGATATGATATGATATGAT 2171
DB 300 AGACAAAGCAGGATCCCTACTTTTGTGAATTCGCTTATGATATGATATGATATGATATGAT 359
QY 2172 ATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCCAGGTTTTCGACCGAACACGTAATGC 2231

```

```

DB 360 ATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTTCACCGAACACGTAATGC 419
QY 2232 TGAGATTGGAAATAGGACATTAAATTCAAACATTTCGAAGAAGCTTTTACATCAGAACA 2291
DB 420 TGAGATTGGAAATAGGACATTAAATTCAAACATTTCGAAGAAGCTTTTACATCAGAACA 479
QY 2292 CTGGCTTGTAGGATATATAAAGTAAAGCACTGTATTAACAGGAGACATTAGATCAAA 2351
DB 480 CTGGCTTGTAGGATATATAAAGTAAAGCACTGTATTAACAGGAGACATTAGATCAAA 539
QY 2352 ACCCTGAGTCACCAACATTTCCTCCAAACAGACAGTATTTCGAAGAAGCTTACCAAG 2411
DB 540 ACCCTGAGTCACCAACATTTCCTCCAAACAGACAGTATTTCGAAGAAGCTTACCAAG 599
QY 2412 GAAGCGTGTCTACATTAATAAATAAGCTGTTTAAAGAAAGCGCAAGAAA 2461
DB 600 GAAGCGTGTCTACATTAATAAATAAGCTGTTTAAAGAAAGCGCAAGAAA 649

RESULT 7
BP871615      582 bp      mRNA      linear      EST 27-JUL-2004
LOCUS      BP871615 Sugano cDNA library, embryonal kidney Homo sapiens cDNA
DEFINITION      clone HKR00887, mRNA sequence.
ACCESSION      BP871615
VERSION      BP871615.1 GI:50716064
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 582)
Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,
Nakai, K. and Sugano S.
Database of Transcriptional Start Sites (DBTSS) for Comparative
Studies of the Promoters of Human and Mouse Genes
Unpublished (2004)
Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp.
Location/Qualifiers
1..582
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HKR00887"
/tissue_type="embryonal kidney"
/cell_line="293"
/clone_lib="Sugano cDNA library, embryonal kidney"

ORIGIN
Query Match      23.5%; Score 582; DB 5; Length 582;
Best Local Similarity 100.0%; Pred. No. 8.1e-302;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1438 TTTGAGCACTATTTGGGGGATGACATGAAGAAGGAAATCCACCTGTGGAGGACAGAGT 1497
DB 1 TTTGAGCACTATTTGGGGGATGACATGAAGAAGGAAATCCACCTGTGGAGGACAGAGT 60
QY 1498 GATGAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGTAAGTGAAGAA 1557
DB 61 GATGAGGATGACAAAGAAACCAAGGAAATTTGTATGATAGGAGGTAAGTGAAGAA 120
QY 1558 CATGCACTGACAGCAAGAAACCTGAGAGGATAGCCCTATATAAAGCAATGTC 1617
DB 121 CATGCACTGACAGCAAGAAACCTGAGAGGATAGCCCTATATAAAGCAATGTC 180
QY 1618 ACCATGTTGATGCTGATGCTATTGATGATGTTTGGTGTCCACTGTACCTGGGTCAAGC 1677
DB 181 ACCATGTTGATGCTGATGCTATTGATGATGTTTGGTGTCCACTGTACCTGGGTCAAGC 240

```



```

|||||
507 TCATACACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCGAGTTTGCACGA 566
|||||
2221 ACAGTAATCTCAGATTTGGAATAGGACATTAATTCARACATTTGGAGAGCCCTT 2280
|||||
567 ACAGTAATCTCAGATTTGGAATAGGACATTAATTCARACATTTGGAGAGCCCTT 626
|||||
2281 ACATCAGAACACTGGCTT 2298
|||||
627 ACATCAGAACACTGGCTT 644

RESULT 3
BX335717 1057 bp mRNA linear EST 08-APR-2004
LOCUS BX335717 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1018YD04 5-PRIME, mRNA sequence.
ACCESSION BX335717
VERSION BX335717.2 GI:46281069
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES 1 (bases 1 to 1057)
JOURNAL Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT On May 2, 2003 this sequence version replaced gi:30345448.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
2459.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0D1018YD04P1&c=2459.f.

FEATURES
source
1..1057
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1018YD04"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 24.5%; Score 608; DB 5; Length 1057;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 720 TGTAAACCTGGTCTGATTTTGGCAATGCTGCTCTATCTCTATCTCTATCTATGCT 779
DB 1 TGTAAACCTGGTCTGATTTTGGCAATGCTGCTCTATCTCTATCTCTATCTATGCT 60
QY 780 CTCTGCTGGGTGGTGTATGATTTATCATCAATCTTATCCACGCGATGATTTGCTT 839
DB 61 CTCTGCTGGGTGGTGTATGATTTATCATCAATCTTATCCACGCGATGATTTGCTT 120
QY 840 GTTACTGATCAGAGATACGAAAGCTCTACATAGCATATAGCACTTCTACATGCT 899
DB 121 GTTACTGATCAGAGATACGAAAGCTCTACATAGCATATAGCACTTCTACATGCT 180
QY 900 GGGTTTAAATATTCATTCAGATACCTTTTGGGATTCAGCCAAATCAGAACAGTGA 959
|||||

```

```

DB 181 GGGTTTAAATATTCATTCAGATACCTTTTGGGATTCAGCCAAATCAGAACAGTGA 240
QY 960 ACACATGGCAGCTGCAGGTGCTTTTGCATTTGCTGCAAGCTTATGCTTTCTTGCAATCT 1019
DB 241 ACACATGGCAGCTGCAGGTGCTTTTGCATTTGCTGCAAGCTTATGCTTTCTTGCAATCT 300
QY 1020 GAGAGACCGATTAAACAAAGAGTTCAGACGCTTTCTTTTGGGTGATCAGTACG 1079
DB 301 GAGAGACCGATTAAACAAAGAGTTCAGACGCTTTCTTTTGGGTGATCAGTACG 360
QY 1080 TGCAGGTGCTGTTCCCTTAGTGTCTATTTTGCATTTATACAGTTTACATTTGCACCATG 1139
DB 361 TGCAGGTGCTGTTCCCTTAGTGTCTATTTTGCATTTATACAGTTTACATTTGCACCATG 420
QY 1140 GAGTGGCAGGTTTATTCATTTGGGATGATCTGGGATGCAAAAATACATTTCCAATAT 1199
DB 421 GAGTGGCAGGTTTATTCATTTGGGATGATCTGGGATGCAAAAATACATTTCCAATAT 480
QY 1200 TGCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGCATCTACATAT 1259
DB 481 TGCATCAGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGCATCTACATAT 540
QY 1260 TCTTGATGATCTTCCAGCAGGCTTTGGTCTGCAATCAAAAATATCAACGATGAAG 1319
DB 541 TCTTGATGATCTTCCAGCAGGCTTTGGTCTGCAATCAAAAATATCAACGATGAAG 600
QY 1320 AGTATTTG 1327
DB 601 AGTATTTG 608

RESULT 4
CR620015 1074 bp mRNA linear HTC 21-JUL-2004
LOCUS CR620015 full-length cDNA clone CS0D1018YD04 of Placenta COT 25-normalized
DEFINITION of Homo sapiens (human).
ACCESSION CR620015
VERSION CR620015.1 GI:50500822
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLES 1 (bases 1 to 1074)
JOURNAL Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
2 (bases 1 to 1074)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
Location/Qualifiers
1..1074
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1018YD04"
/tissue_type="Placenta COT 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 24.5%; Score 608; DB 3; Length 1074;
Best Local Similarity 100.0%; Pred. No. 0;

```

1691 GTCCAGGTGATGCTCTGGCTCATACATATCATGATGCGCAGGAAATATCTTAGATGATT 1750
 Db GTCCAGGTGATGCTCTGGCTCATACATATCATGATGCGCAGGAAATATCTTAGATGATT 120
 1751 TTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATGACAGGAAATGCTT 1810
 Db TTAGAGAGCTTACTTTTGGCTAGGCAAAATACAGATGACATGACAGGAAATGCTT 180
 1811 GGTGGATATGGTATCAGATAGCTGGAATGCTTAATAGAACTAGCTTGGTGGATAATA 1870
 Db GGTGGATATGGTATCAGATAGCTGGAATGCTTAATAGAACTAGCTTGGTGGATAATA 240
 1871 ACACCTGGAATAACAGCCACATACATGCTGGTGGAAAGCTATGCTTCTTAATGAACAG 1930
 Db ACACCTGGAATAACAGCCACATACATGCTGGTGGAAAGCTATGCTTCTTAATGAACAG 300
 1931 CAGCCTATAAATCATGAGGACTCTAGATAGATATGTTTGGTATATTTTGGAGGG 1990
 Db CAGCCTATAAATCATGAGGACTCTAGATAGATATGTTTGGTATATTTTGGAGGG 360
 1991 TTATGGCTATCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAG 2050
 Db TTATGGCTATCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAG 420
 2051 GAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGAGAAATCCGTTG 2110
 Db GAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGAGAAATCCGTTG 480
 2111 TAGACAAAGCAGGATCCCTACTTTTGTGAATGCTTATGATAAATGTCTACTACA 2170
 Db TAGACAAAGCAGGATCCCTACTTTTGTGAATGCTTATGATAAATGTCTACTACA 540
 2171 GATTGGAGAAATCAGCTGATATTCGTACACCCCGAGTTTGGACGAAACAGCTAATG 2230
 Db GATTGGAGAAATCAGCTGATATTCGTACACCCCGAGTTTGGACGAAACAGCTAATG 600
 2231 CTGAGATTGGAATAAGACATTAATTAATTTGGAAGAGCTTTTACATCAGAAC 2290
 Db CTGAGATTGGAATAAGACATTAATTAATTTGGAAGAGCTTTTACATCAGAAC 660
 2291 ACTGGCTTTAGGATATATAAGTAAAGCACTGATACAGGAGAGACATTAGATCACA 2350
 Db ACTGGCTTTAGGATATATAAGTAAAGCACTGATACAGGAGAGACATTAGATCACA 720
 2351 AACCTCGAGTCACCAACATTTCCCAAAACAGAGATTTTGTCAAAGAGAGCTACCAAAA 2410
 Db AACCTCGAGTCACCAACATTTCCCAAAACAGAGATTTTGTCAAAGAGAGCTACCAAAA 780
 2411 GGAAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAAATATCTAAGA 2470
 Db GGAAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAAATATCTAAGA 840
 2471 AGACTGTTTAA 2481
 Db AGACTGTTTAA 851
 RESULT 2
 CF995035
 LOCUS
 DEFINITION AGENCOURT 15622198 NIH MGC 147 Homo sapiens cdna clone
 IMAGE:30528919 5', mRNA sequence.
 CF995035
 ACCESSION CF995035.1 GI:38511095
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 861)
 NIH-MGC <http://mgi.nci.nih.gov/>
 AUTHORS
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: gcapbs@mail.nih.gov
 Tissue Procurement: Dr. Stefan Hansson
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
 and advice from Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: MDAMB15 row: k column: 08
 High quality sequence stop: 625.

FEATURES
source

1. 861
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30528919"
 /tissue_type="Human Placenta"
 /lab_host="DH10B Tona"
 /clone_lib="NIH MGC 147"
 /notes="Organ: placenta; Vector: pBluescriptR; Site 1:
 alt-XhoI; Site 2: BamH; Oligo-dT primed using primer
 5'-TTTTTTTTTTTTTTVN-3', size-selected for average
 insert size 2.3 kb and normalized to 50. This is a
 primary library enriched for full-length clones and
 constructed using the Cap-trapper method (Carninci, in
 preparation). Library constructed by M. Brownstein
 (NIH/NHGRI, National Institutes of Health). Note: This is
 a NIH_MGC library."

ORIGIN

Query Match 24.9%; Score 618; DB 7; Length 861;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| Qy | 1681 | GCCTACTCTAGTCCAAAGTCTAGTCTCTGGCTCATACATATCATGATGCGCAGGAAATATC | 1740 |
| Db | 27 | GCCTACTCTAGTCCAAAGTCTAGTCTCTGGCTCATACATATCATGATGCGCAGGAAATATC | 86 |
| Qy | 1741 | TTAGATGATTTTGAAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATCCACGA | 1800 |
| Db | 87 | TTAGATGATTTTGAAGAGCTTACTTTTGGCTAAGCAAAATACAGATGAACATCCACGA | 146 |
| Qy | 1801 | GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAAATGGCTAATAGAACTACGTTG | 1860 |
| Db | 147 | GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAAATGGCTAATAGAACTACGTTG | 206 |
| Qy | 1861 | GTGGATAATAACACCTGGAAATACACCCACATAGCACTGGTGGGAAAGCTATGCTTCT | 1920 |
| Db | 207 | GTGGATAATAACACCTGGAAATACACCCACATAGCACTGGTGGGAAAGCTATGCTTCT | 266 |
| Qy | 1921 | AATGAAACAGCAGCCTATAAAATCATAGGACTCTAGATAGTATGTTTGGTTAT | 1980 |
| Db | 267 | AATGAAACAGCAGCCTATAAAATCATAGGACTCTAGATAGTATGTTTGGTTAT | 326 |
| Qy | 1981 | TTTGGAGGGGTTATTTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGG | 2040 |
| Db | 327 | TTTGGAGGGGTTATTTGGCTATTTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAGG | 386 |
| Qy | 2041 | ATAGCTGAAGAGAGAACATCCCAAGACATTCGGGAAAGTGAATAATTTACCCCAAGGGA | 2100 |
| Db | 387 | ATAGCTGAAGAGAGAACATCCCAAGACATTCGGGAAAGTGAATAATTTACCCCAAGGGA | 446 |
| Qy | 2101 | GAATTCCTGTAGACAAACAGCAGTCCCTACTTTGTTGAATGCTTATGTAATAAATG | 2160 |
| Db | 447 | GAATTCCTGTAGACAAACAGCAGTCCCTACTTTGTTGAATGCTTATGTAATAAATG | 506 |
| Qy | 2161 | TCATATCTACAGATTTTGGAGAAATGCGAGCTGGATTTTCGTACACCCCAAGGTTTGGACCGA | 2220 |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:16:54 ; Search time 11387.4 Seconds
(without alignments)
7939.229 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcgagccctcgccccc.....tatctaagaagacttttaa 2481

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 32822875 seqs, 1821985908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hsc.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_gsl1.*

9: gb_gsl2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 851 | 34.3 | 2212 | 3 | CR610915 full-length |
| 2 | 618 | 24.9 | 861 | 7 | CF995035 AGENCOURT |
| 3 | 608 | 24.5 | 1057 | 5 | EX335717 BX335717 |
| 4 | 608 | 24.5 | 1074 | 3 | CR620015 full-length |
| 5 | 598 | 24.1 | 709 | 7 | CF780700 AGENCOURT |
| 6 | 590 | 23.8 | 903 | 6 | CD358477 AGENCOURT |
| 7 | 582 | 23.5 | 592 | 5 | BP871615 BP871615 |
| 8 | 582 | 23.5 | 589 | 5 | BQ877881 AGENCOURT |
| 9 | 562 | 22.7 | 562 | 6 | CE128931 K-EST0178 |
| 10 | 549 | 22.1 | 735 | 4 | BG698273 602860188 |
| 11 | 546 | 22.0 | 627 | 4 | BG434322 602506307 |
| 12 | 542 | 21.8 | 577 | 7 | CN263069 170006001 |
| 13 | 522 | 21.0 | 778 | 4 | BG573592 602594608 |
| 14 | 517 | 20.8 | 776 | 4 | BG620025 602618240 |
| 15 | 508 | 20.5 | 793 | 6 | CD103829 AGENCOURT |
| 16 | 499 | 20.1 | 691 | 4 | BM785621 K-EST0064 |
| 17 | 495 | 20.0 | 1088 | 5 | EX332111 BX332111 |
| 18 | 487 | 19.6 | 689 | 2 | BF970612 602773941 |
| 19 | 478 | 19.3 | 521 | 1 | AA307802 EST178658 |
| 20 | 465 | 18.7 | 876 | 5 | BUI17067 AGENCOURT |
| 21 | 464 | 18.7 | 515 | 7 | CR540551 DKFZP459M |
| 22 | 456 | 18.4 | 594 | 1 | BM745603 K-EST0019 |
| 23 | 448 | 18.1 | 785 | 6 | CB963947 AGENCOURT |
| 24 | 435 | 17.5 | 1082 | 2 | BF981298 602308544 |

| | | | | | |
|----|-----|------|-----|---|-----------|
| 25 | 434 | 17.5 | 972 | 4 | BG249115 |
| 26 | 423 | 17.0 | 730 | 4 | BG105696 |
| 27 | 422 | 17.0 | 556 | 4 | BM749457 |
| 28 | 421 | 17.0 | 696 | 7 | CN263068 |
| 29 | 416 | 16.8 | 855 | 5 | BUI940129 |
| 30 | 389 | 15.7 | 510 | 4 | BG396740 |
| 31 | 381 | 15.4 | 916 | 5 | BU902762 |
| 32 | 380 | 15.3 | 576 | 1 | AL598898 |
| 33 | 380 | 15.3 | 914 | 5 | BUI151612 |
| 34 | 372 | 15.0 | 879 | 5 | BQ425152 |
| 35 | 366 | 14.8 | 702 | 4 | BI254237 |
| 36 | 365 | 14.7 | 703 | 1 | AI743228 |
| 37 | 364 | 14.7 | 713 | 1 | AI492174 |
| 38 | 345 | 13.9 | 869 | 5 | BQ422643 |
| 39 | 333 | 13.4 | 757 | 4 | BI69525 |
| 40 | 329 | 13.3 | 395 | 2 | BE167429 |
| 41 | 327 | 13.2 | 517 | 5 | BU683595 |
| 42 | 319 | 12.9 | 629 | 5 | BQ786822 |
| 43 | 318 | 12.8 | 485 | 1 | AI470192 |
| 44 | 317 | 12.8 | 527 | 5 | BU934287 |
| 45 | 315 | 12.7 | 542 | 1 | AI355318 |

ALIGNMENTS

RESULT 1

CR610915

LOCUS

DEFINITION

full-length cDNA clone CSODK010Y001 of HeLa cells

of Homo sapiens (human)

ACCESSION CR610915

VERSION CR610915.1

KEYWORDS HTC; CNSLT_CDNA

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2212)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/

REFERENCE 2 (bases 1 to 2212)

AUTHORS Genoscope

TITLE Direct Submission

JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequenage

COMMENT BP 131 31006 ERY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

FEATURES

1..2212

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODK010Y001"

/tissue_type="HeLa cells"

/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 34.3%; Score 851; DB 3; Length 2212;

Best Local Similarity 100.0%; Fred. No. 0;

Matches 851; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1631 TGATGCTATTGATGATGTTTGTCTGCTACCTGCTCAGCAAGCAATCCCTACTCTA 1690

Db 1 TGATGCTATTGATGATGTTTGTCTGCTACCTGCTCAGCAAGCAATCCCTACTCTA 60

Db 181 GACTCTAGATGTAGATTATGTTTGGTTATTTTGGAGGGGTTATTGGCTATTCTCGTGA 240
Qy 2010 TGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGGAGAACATCCCAAGACAT 2069
Db 241 TGATATCAACAAATTTCTCTGGATGGTTAGATAGCTGAAGGAGAACATCCCAAGACAT 300
Qy 2070 TCGGAAAGTGACTATTTTACCCACACAGGAGAAATCCGTGTAGACAAAGCAGGATCCCC 2129
Db 301 TCGGAAAGTGACTATTTTACCCACACAGGAGAAATCCGTGTAGACAAAGCAGGATCCCC 360
Qy 2130 TACTTTGTGAATGGCTTATGTATATAAATGTCTACTACAGATTGGAGAAATGCAGCT 2189
Db 361 TACTTTGTGAATGGCTTATGTATATAAATGTCTACTACAGATTGGAGAAATGCAGCT 420
Qy 2190 GGATTTTCGT 2199
Db 421 GGATTTTCGT 430

Search completed: December 16, 2004, 16:17:34
Job time : 2198.32 secs

```
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-878-178-717

Query Match      17.3%; Score 430; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e-216;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1770 GCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 1829
Db 1 GCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 60

QY 1830 GATAGCTGGATGGCTAATAGAACTACGTTGGTGGGATAATAACACCTGGAATAACAGCCA 1889
Db 61 GATAGCTGGATGGCTAATAGAACTACGTTGGTGGGATAATAACACCTGGAATAACAGCCA 120

QY 1890 CATAGCACTGGTGGGAAAGCTATGCTCTTAATAAGAAACAGACGACCTATAAAATCATGAG 1949
Db 121 CATAGCACTGGTGGGAAAGCTATGCTCTTAATAAGAAACAGACGACCTATAAAATCATGAG 180

QY 1950 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGGTGA 2009
Db 181 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGGTGA 240

QY 2010 TGATATCAACAAATTTCTTGGATGTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 2069
Db 241 TGATATCAACAAATTTCTTGGATGTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 300

QY 2070 TCGGGAAGTGAATTTTACCCACAGGAGAAATCCCGTAGACAAAGCAGGATCCCC 2129
Db 301 TCGGGAAGTGAATTTTACCCACAGGAGAAATCCCGTAGACAAAGCAGGATCCCC 360

QY 2130 TACTTTGTTGAATTCCTTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 2189
Db 361 TACTTTGTTGAATTCCTTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 420

QY 2190 GGATTTTCGT 2199
Db 421 GGATTTTCGT 430

RESULT 14
US-10-046-935-717
; Sequence 717, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-717

Query Match      17.3%; Score 430; DB 13; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e-216;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1770 GCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 1829
```

```
Db 1 GCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 60

QY 1830 GATAGCTGGATGGCTAATAGAACTACGTTGGTGGGATAATAACACCTGGAATAACAGCCA 1889
Db 61 GATAGCTGGATGGCTAATAGAACTACGTTGGTGGGATAATAACACCTGGAATAACAGCCA 120

QY 1890 CATAGCACTGGTGGGAAAGCTATGCTCTTAATAAGAAACAGACGACCTATAAAATCATGAG 1949
Db 121 CATAGCACTGGTGGGAAAGCTATGCTCTTAATAAGAAACAGACGACCTATAAAATCATGAG 180

QY 1950 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGGTGA 2009
Db 181 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGGTGA 240

QY 2010 TGATATCAACAAATTTCTTGGATGTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 2069
Db 241 TGATATCAACAAATTTCTTGGATGTTAGGATAGCTGAAGGAGAAACATCCCAAAGACAT 300

QY 2070 TCGGGAAGTGAATTTTACCCACAGGAGAAATCCCGTAGACAAAGCAGGATCCCC 2129
Db 301 TCGGGAAGTGAATTTTACCCACAGGAGAAATCCCGTAGACAAAGCAGGATCCCC 360

QY 2130 TACTTTGTTGAATTCCTTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 2189
Db 361 TACTTTGTTGAATTCCTTATGTAATAAATGTCATACAGATTTGGAGAAATGCAGCT 420

QY 2190 GGATTTTCGT 2199
Db 421 GGATTTTCGT 430

RESULT 15
US-10-146-502-717
; Sequence 717, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-146-502-717

Query Match      17.3%; Score 430; DB 14; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.4e-216;
Matches 430; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1770 GCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 1829
Db 1 GCTAAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGGTGGGATATGGCTATCA 60

QY 1830 GATAGCTGGATGGCTAATAGAACTACGTTGGTGGGATAATAACACCTGGAATAACAGCCA 1889
Db 61 GATAGCTGGATGGCTAATAGAACTACGTTGGTGGGATAATAACACCTGGAATAACAGCCA 120

QY 1890 CATAGCACTGGTGGGAAAGCTATGCTCTTAATAAGAAACAGACGACCTATAAAATCATGAG 1949
Db 121 CATAGCACTGGTGGGAAAGCTATGCTCTTAATAAGAAACAGACGACCTATAAAATCATGAG 180

QY 1950 GACTCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTATTTGGCTATTTCTGGTGA 2009
```

RESULT 11

US-10-296-115-629
; Sequence 629, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; PRIORITY FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIORITY FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIORITY FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 629
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-629

Query Match 26.3%; Score 653; DB 16; Length 1114;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 177 CGGGGGCTGTGCGAGCCGGCTGGGTGGGAGTGGTCTCTCTCCATCACCATCCTCTTCT 236
DB 388 CGGGGGCTGTGCGAGCCGGCTGGGTGGGAGTGGTCTCTCTCTCTCTCTCTCTCTCT 447
QY 237 GGCTCGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTATCCGCTTCGGAAGCATCAT 296
DB 448 GGCTCGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTATCCGCTTCGGAAGCATCAT 507
QY 297 CCACGAGTTCGACCCGCTGGTTAACTATATAGATCAACATCATCTTGCATCTCATGGGT 356
DB 508 CCACGAGTTCGACCCGCTGGTTAACTATATAGATCAACATCATCTTGCATCTCATGGGT 567
QY 357 CTATGAATTTTAAATGGTTTGAATGAAGAGCATGGTATCCATAGAAAGATAGTAGG 416
DB 568 CTATGAATTTTAAATGGTTTGAATGAAGAGCATGGTATCCATAGAAAGATAGTAGG 627
QY 417 TGCTACTGTTTACCAGGGTGTGATGATACCCGCTGGCTTATTCATGGATTTTAAATAC 476
DB 628 TGCTACTGTTTACCAGGGTGTGATGATACCCGCTGGCTTATTCATGGATTTTAAATAC 687
QY 477 ATGACATATGTTTCATATAGAGAGCATGTATGTGTCTTCCGCTTATTCATGGATTTTAAATAC 536
DB 688 ATGACATATGTTTCATATAGAGAGCATGTATGTGTCTTCCGCTTATTCATGGATTTTAAATAC 747
QY 537 CCTTACATCTATATCTACTTCTCTTACAGAGAACTTTGGAAACCAAGAGCAGGACT 596
DB 748 CCTTACATCTATATCTACTTCTCTTACAGAGAACTTTGGAAACCAAGAGCAGGACT 807
QY 597 TTTAGTCTGTTTATTTATGTTATGTTACAGGCTACATATCTCGGTCAGTGGATC 656
DB 808 TTTAGTCTGTTTATTTATGTTATGTTACAGGCTACATATCTCGGTCAGTGGATC 867
QY 657 CTTTCATATGAAGCATGTCTATTTTGGCACTTCAGTTCATATCTTATTTATGGTAAA 716
DB 868 CTTTCATATGAAGCATGTCTATTTTGGCACTTCAGTTCATATCTTATTTATGGTAAA 927
QY 717 ATCTGTAAATCGGTCAGTTTGTGGCAATGTGCTGCTTATCTATCTTATCTATAT 776
DB 928 ATCTGTAAATCGGTCAGTTTGTGGCAATGTGCTGCTTATCTATCTTATCTATAT 987
QY 777 GGTCTCTGTTGGGTGGTATGATTTATCATCAATCTTATTCACCTGCATG 829
DB 988 GGTCTCTGTTGGGTGGTATGATTTATCATCAATCTTATTCACCTGCATG 1040

RESULT 12

US-09-998-598-1643
; Sequence 1643, Application US/09998598

Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; PRIORITY FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1643
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1643

Query Match 20.2%; Score 500; DB 9; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.1e-253;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 471 AAATACATTCGAACATACTCTTACATAGAGAGTATGTGTCTCTTGCACCACTTT 530
DB 1 AAATACATTCGAACATACTCTTACATAGAGAGTATGTGTCTCTTGCACCACTTT 60
QY 531 TAGCGGCTTACATCTATATCTACTTCTCTCTTACAGAGAACTTTGGAACCAAGAGC 590
DB 61 TAGCGGCTTACATCTATATCTACTTCTCTCTTACAGAGAACTTTGGAACCAAGAGC 120
QY 591 AGGACTTTTACGCTGCTTTTATTTGCTATTTGACAGGCTACATATCTCGGTCACTAGC 650
DB 121 AGGACTTTTACGCTGCTTTTATTTGCTATTTGACAGGCTACATATCTCGGTCACTAGC 180
QY 651 TGAATCTTTGATTAAGAGCATTTGCTATTTTGGCACTTCACTTCACTATCTATTTATG 710
DB 181 TGAATCTTTGATTAAGAGCATTTGCTATTTTGGCACTTCACTTCACTATCTATTTATG 240
QY 711 GGTAAATCTGTAATAAAGTGGTCACTTTTGGCACTTCACTTCACTATCTATTTATG 770
DB 241 GGTAAATCTGTAATAAAGTGGTCACTTTTGGCACTTCACTTCACTATCTATTTATG 300
QY 771 CTATATGCTCTGCTGGGTGGTATGTTATTTATCATCTTATTTCCACTGCATGT 830
DB 301 CTATATGCTCTGCTGGGTGGTATGTTATTTATCATCTTATTTCCACTGCATGT 360
QY 831 ATTGTGTTGTTACTGATGAGATACAGCAAGAGTCTACATAGCATATAGCACTTT 890
DB 361 ATTGTGTTGTTACTGATGAGATACAGCAAGAGTCTACATAGCATATAGCACTTT 420
QY 891 CTACATTTGGGTTTAAATTTATTTATCATGAGATACCTTTTGGGATTCAGCAATCAG 950
DB 421 CTACATTTGGGTTTAAATTTATTTATCATGAGATACCTTTTGGGATTCAGCAATCAG 480
QY 951 AACAGTGAACATCGGAG 970
DB 481 AACAGTGAACATCGGAG 500

RESULT 13

US-09-878-178-717
; Sequence 717, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08


```

Db 901 TACTGGGTATGCAAAATACACATTCCTCAATTAATGTCATCAGTGTCTGAGCATCAACCTAC 960
Qy 1227 GACTTGGGTCTCTTCTCTTCTGATACATATCTTGTATGATCTTCCAGAGGCGCT 1286
Db 961 GACTTGGGTCTCTTCTCTTCTGATACATATCTTGTATGATCTTCCAGAGGCGCT 1020
Qy 1287 TTGGTCTGTCATCAAAATATCAACGATGAAGA 1320
Db 1021 TTGGTCTGTCATCAAAATATCAACGATGAAGA 1054

RESULT 7
US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US20030211472A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Query Match 42.5%; Score 1054; DB 10; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 CTTCGGCGTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCGTGGTTAACTATAG 325
Db 1 CTTCGGCGTCATCCGCTTCGAAGCATCATCCAGAGTTCGACCGTGGTTAACTATAG 60
Qy 327 ATCAACACATCATCTTGCATCTCATGGGTTCTATGAATTTTAAATGGTTGATGAAG 386
Db 61 ATCAACACATCATCTTGCATCTCATGGGTTCTATGAATTTTAAATGGTTGATGAAG 120
Qy 387 AGCATGGTATCCACTAGGAAGATAGTAGGTGTACTGTATACCCAGGTTGATGATAAC 446
Db 121 AGCATGGTATCCACTAGGAAGATAGTAGGTGTACTGTATACCCAGGTTGATGATAAC 180
Qy 447 CGTGGCGTTATTCATTTGATGATTTTAAATCATTTGAACATACTGTTCACATAAGAGAGCT 506
Db 181 CGTGGCGTTATTCATTTGATGATTTTAAATCATTTGAACATACTGTTCACATAAGAGAGCT 240
Qy 507 ATGTGTGTTCTTGCACCACTTTTAGCGGCTTACATCTATATCTTCTCTGCTTAC 566
Db 241 ATGTGTGTTCTTGCACCACTTTTAGCGGCTTACATCTATATCTTCTCTGCTTAC 300
Qy 567 AAGAGAACTTTGAAACCAAGGACGAGGACTTTTAGCTGTTGTTTATGATGATGACC 626
Db 301 AAGAGAACTTTGAAACCAAGGACGAGGACTTTTAGCTGTTGTTTATGATGATGACC 360
Qy 627 AGGCTACATATCTCGGTGAGTGTGATCTTTCATATGAGGATGATGATGATGATG 686
Db 361 AGGCTACATATCTCGGTGAGTGTGATCTTTCATATGAGGATGATGATGATGATG 420
Qy 687 ACTTCAGTTCACATATCTTTTATGGTAAATCTGTAAATCTGGTGGTGGTGGTGGTGG 746
Db 421 ACTTCAGTTCACATATCTTTTATGGTAAATCTGTAAATCTGGTGGTGGTGGTGGTGG 480

```

```

Qy 747 AATGTGCTGCTCTTATCTATTTCTATGCTCTCTGCTGGGTGCTTATGATTTAT 806
Db 481 AATGTGCTGCTCTTATCTATTTCTATGCTCTCTGCTGGGTGCTTATGATTTAT 540
Qy 807 CATCAATCTTATCCACATGCTATTTGTTGTTACTGATGAGAGATACAGCAAAAG 866
Db 541 CATCAATCTTATCCACATGCTATTTGTTGTTACTGATGAGAGATACAGCAAAAG 600
Qy 867 AGTCTACATAGCATATAGCAGCTTTCTACATTTGGGTTTAAATATTAATCAATGAGATPACC 926
Db 601 AGTCTACATAGCATATAGCAGCTTTCTACATTTGGGTTTAAATATTAATCAATGAGATPACC 660
Qy 927 TTTTGTGGGATCCAGCCCAATCAGAAAGTGAACACATGCGAGCTGCAGGTGCTTTTC 986
Db 561 TTTTGTGGGATTCAGCCCAATCAGAAAGTGAACACATGCGAGCTGCAGGTGCTTTTC 720
Qy 987 ATTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAAACAAAACAGAGTT 1046
Db 721 ATTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAAACAAAACAGAGTT 780
Qy 1047 CCAGACCCCTTTCTTTTGGGTTGATCACTAGCTGCGAGGTGCTGTTCTTATGTCAT 1106
Db 781 CCAGACCCCTTTCTTTTGGGTTGATCACTAGCTGCGAGGTGCTGTTCTTATGTCAT 840
Qy 1107 CTATTGTGATATACAGGTTACATTTGCACCATGGAGTGGCAGGTTTATTCATTGTGGGA 1166
Db 841 CTATTGTGATATACAGGTTACATTTGCACCATGGAGTGGCAGGTTTATTCATTGTGGGA 900
Qy 1167 TACTGGGTATGCAAAATACACATTCATTTGATGATGCTGCTGAGCATCAACCTAC 1226
Db 901 TACTGGGTATGCAAAATACACATTCATTTGATGATGCTGCTGAGCATCAACCTAC 960
Qy 1227 GACTTGGGTGCTTTCTTCTTGTGATACATATTTGATGATGCTTCCAGAGGCT 1286
Db 961 GACTTGGGTGCTTTCTTCTTGTGATACATATTTGATGATGCTTCCAGAGGCT 1020
Qy 1287 TTGGTCTGTCATCAAAATATCAACGATGAAGA 1320
Db 1021 TTGGTCTGTCATCAAAATATCAACGATGAAGA 1054

RESULT 8
US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US20040185440A9
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Query Match 42.5%; Score 1054; DB 11; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


Db 601 AGTCTACATAGCATATAGCATTCTACATTGTGGTTTAATATATCAATGCAGATACC 660
Qy 927 TTTTGTGGATCCAGCCAAATCAGAACAGTGAACACATGCGAGTGTCTTTGC 986
Db 661 TTTTGTGGATCCAGCCAAATCAGAACAGTGAACACATGCGAGTGTCTTTGC 720
Qy 987 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCCGATTAACAAACAAAGATT 1046
Db 721 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGACCCGATTAACAAACAAAGATT 780
Qy 1047 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTCAGGTGCTGTCTTCTAGTGAT 1106
Db 781 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTCAGGTGCTGTCTTCTAGTGAT 840
Qy 1107 CTATTGACTTATACAGGTTCATATCCACCATGAGTGCAGGTTTATTCAATGTGGGA 1166
Db 841 CTATTGACTTATACAGGTTCATATGCACCATGAGTGCAGGTTTATTCAATGTGGGA 900
Qy 1167 TACTGGGTATGCAAAATATACATTCCTCAATTTATTCATCAGTGTCTGAGCATCAACCTAC 1226
Db 901 TACTGGGTATGCAAAATATACATTCCTCAATTTATTCATCAGTGTCTGAGCATCAACCTAC 960
Qy 1227 GACTTGGGTGCTTTCTTTTGGGTGATCACTAGCTCAGGTGCTGTCTTCTAGTGAT 1286
Db 961 GACTTGGGTGCTTTCTTTTGGGTGATCACTAGCTCAGGTGCTGTCTTCTAGTGAT 1020
Qy 1287 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1320
Db 1021 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1054

RESULT 5

US-09-974-879-133
; Sequence 133, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066,094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,089

Query Match 42.5%; Score 1054; DB 10; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 CTTTCGGCGTCAATCGGCTTCGAAAGCATCATCCAGAGTTCGACCGTGGTTAACTATAG 326
Db 1 CTTTCGGCGTCAATCGGCTTCGAAAGCATCATCCAGAGTTCGACCGTGGTTAACTATAG 60
Qy 327 ATCAACACATCATCTTGCATCTCATGGGTTCATGAATTTTAAATGGTTGATGAAG 386
Db 61 ATCAACACATCATCTTGCATCTCATGGGTTCATGAATTTTAAATGGTTGATGAAG 120
Qy 387 AGCATGTATCCACTAGGAAGATAGTAGTGTCTGTCTTACCCAGGGTTCATGATAAC 446
Db 121 AGCATGTATCCACTAGGAAGATAGTAGTGTCTGTCTTACCCAGGGTTCATGATAAC 180
Qy 447 CGCTGGCCTTATTTCATTGGATTTTAAATACATTTGAACATACTGTTTCATAGAGAGCT 506
Db 181 CGCTGGCCTTATTTCATTGGATTTTAAATACATTTGAACATACTGTTTCATAGAGAGCT 240
Qy 507 ATGTGTGTTCTTGCACCAACTTTAGCGGCTTACATCTATATCTACTTTCTCTGTTAC 566
Db 241 ATGTGTGTTCTTGCACCAACTTTAGCGGCTTACATCTATATCTACTTTCTCTGTTAC 300
Qy 567 AAGAGAACTTTTGAACCAAGGAGCAGGACTTTTAGCTGTCTTGTATTGCTATTGTACC 626
Db 301 AAGAGAACTTTTGAACCAAGGAGCAGGACTTTTAGCTGTCTTGTATTGCTATTGTACC 360
Qy 627 AGGCTACATATCTCGGTGAGTAGCTGATCCTTTGATATGAAGGCAATGCTATTGTC 686
Db 361 AGGCTACATATCTCGGTGAGTAGCTGATCCTTTGATATGAAGGCAATGCTATTGTC 420
Qy 687 ACTTCAGTTCACATATCTTTATGGGTGATTAATCTGTAATAAATCTGGGTGCTAGTTTGGAC 746
Db 421 ACTTCAGTTCACATATCTTTATGGGTGATTAATCTGTAATAAATCTGGGTGCTAGTTTGGAC 480
Qy 747 AATGTGCTGTCTTATCTTCTTATATGTTCTCTGCTGCTGGGTGTTGTTATTTAT 806
Db 481 AATGTGCTGTCTTATCTTCTTATATGTTCTCTGCTGCTGGGTGTTGTTATTTAT 540
Qy 807 CATCAATCTTTATCCACTGATGTTTGTGTTGTTTACTGATGAGAGATACAGCAAAAG 866
Db 541 CATCAATCTTTATCCACTGATGTTTGTGTTGTTTACTGATGAGAGATACAGCAAAAG 600
Qy 867 AGTCTACATAGCATATAGCACTTTCTACATTTGTTGGTTTAAATATTAATGAGATACC 926
Db 601 AGTCTACATAGCATATAGCACTTTCTACATTTGTTGGTTTAAATATTAATGAGATACC 660
Qy 927 TTTTGTGGGATTCAGCCAAATCAGAACAGTGAACATGGGAGCTGAGGTGCTTTGC 986
Db 661 TTTTGTGGGATTCAGCCAAATCAGAACAGTGAACATGGGAGCTGAGGTGCTTTGC 720
Qy 987 ATTGCTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGGATTAACAAACAAAGATT 1046

Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CTTGCGCGTCATCGGCTTCGAAAGCATCATCCAGAGTTCCGACCGGTGTTAACTATAG 326
 Db 1 CTTGCGCGTCATCGGCTTCGAAAGCATCATCCAGAGTTCCGACCGGTGTTAACTATAG 60

QY 327 ATCAACACATCATCTTGGCATCTCATGGTCTTANGAATTTTAAATTTGGTTTGATGAAG 386
 Db 61 ATCAACACATCATCTTGGCATCTCATGGTCTTANGAATTTTAAATTTGGTTTGATGAAG 120

QY 387 AGCATGGTATCCACTAGGAAGATAGTAGGTGCTACTGTTTACCCAGGGTTCATGATATAC 446
 Db 121 AGCATGGTATCCACTAGGAAGATAGTAGGTGCTACTGTTTACCCAGGGTTCATGATATAC 180

QY 447 CGTGCGCTTATTCATTTGATGATTTTAAATACATTAAGTAACTGTTTACATGAAGAGCT 506
 Db 181 CGTGCGCTTATTCATTTGATGATTTTAAATACATTAAGTAACTGTTTACATGAAGAGCT 240

QY 507 ATGTGTGTTCTTTCGACCACTTTTACCGGCTTACATCTATATCTACTTCTCTCTTAC 566
 Db 241 ATGTGTGTTCTTTCGACCACTTTTACCGGCTTACATCTATATCTACTTCTCTCTTAC 300

QY 567 AGAGAACTTTGGAACCAAGGAGCAGGACTTTTAGCTGCTTGTGTTTATTTGCTATTGTACC 626
 Db 301 AGAGAACTTTGGAACCAAGGAGCAGGACTTTTAGCTGCTTGTGTTTATTTGCTATTGTACC 360

QY 627 AGGCTACATATCTCGGTCTAGTAGTGGATCCCTTGTGATGAAGGATGCTATTGTTGC 686
 Db 361 AGGCTACATATCTCGGTCTAGTAGTGGATCCCTTGTGATGAAGGATGCTATTGTTGC 420

QY 687 ACTTCAGTTTACATATCTTTTATGGTAAATCTGTAAATCTGGTTCAGTTTGTGAC 746
 Db 421 ACTTCAGTTTACATATCTTTTATGGTAAATCTGTAAATCTGGTTCAGTTTGTGAC 480

QY 747 AATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGGGTGTTATGTTAT 806
 Db 481 AATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGGGTGTTATGTTAT 540

QY 807 CATCAATCTTATTCCTACTGCATGTTATTTGTTGTTTACTGTATGAGAGATACAGAAAG 866
 Db 541 CATCAATCTTATTCCTACTGCATGTTATTTGTTGTTTACTGTATGAGAGATACAGAAAG 600

QY 867 AGTCTACATAGCATATGACATTTCTACATTTGGGTTTAAATTTATCAATGACATACC 926
 Db 601 AGTCTACATAGCATATGACATTTCTACATTTGGGTTTAAATTTATCAATGACATACC 660

QY 927 TTTTGGGATTTCCAGCCAAATCAGACAGTGAACACATGGCAGCTGAGGTGCTTTGC 986
 Db 661 TTTTGGGATTTCCAGCCAAATCAGACAGTGAACACATGGCAGCTGAGGTGCTTTGC 720

QY 987 ATTGTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAAACAAACAGAGTT 1046
 Db 721 ATTGTGCAAGCTTATGCTTTCTTGCAGTATCTGAGAGACCGATTAAACAAACAGAGTT 780

QY 1047 CCAGACCTTTCTTTTGGGTGATCACTAGCTGAGGTGCTGTTCTTGTAGTGCTAT 1106
 Db 781 CCAGACCTTTCTTTTGGGTGATCACTAGCTGAGGTGCTGTTCTTGTAGTGCTAT 840

QY 1107 CTATTGACTTATACAGGTTTACATTCACCACTGAGTGGCAGGTTTATTCATTGTGGA 1166
 Db 841 CTATTGACTTATACAGGTTTACATTCACCACTGAGTGGCAGGTTTATTCATTGTGGA 900

QY 1167 TACTGGGTATGAAAAATACACATTTCCAAATTTTGCATCAGTGTCTGAGCATCAACCTAC 1226
 Db 901 TACTGGGTATGAAAAATACACATTTCCAAATTTTGCATCAGTGTCTGAGCATCAACCTAC 960

QY 1227 GACTTGGGTGCTTTCTTTCTTTGATCTACATATTTCTGATGATCTTCCAGAGGCT 1286
 Db 961 GACTTGGGTGCTTTCTTTCTTTGATCTACATATTTCTGATGATCTTCCAGAGGCT 1020

QY 1287 TTGGTTCTGCATCAAAAATATCAACGATGAAGA 1320
 Db 1021 TTGGTTCTGCATCAAAAATATCAACGATGAAGA 1054

RESULT 4
 US-10-264-237-412
 ; Sequence 412, Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birst et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P131P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; PRIOR FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 412
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1203)..(1203)
 ; OTHER INFORMATION: n equals, a,t,g, or c
 US-10-264-237-412

Query Match : 42.5%; Score 1054; DB 16; Length 1209;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CTTGCGCGTCATCGGCTTCGAAAGCATCATCCAGAGTTCCGACCGGTGTTAACTATAG 326
 Db 1 CTTGCGCGTCATCGGCTTCGAAAGCATCATCCAGAGTTCCGACCGGTGTTAACTATAG 60

QY 327 ATCAACACATCATCTTGGCATCTCATGGTCTTANGAATTTTAAATTTGGTTTGATGAAG 386
 Db 61 ATCAACACATCATCTTGGCATCTCATGGTCTTANGAATTTTAAATTTGGTTTGATGAAG 120

QY 387 AGCATGGTATCCACTAGGAAGATAGTAGGTGCTACTGTTTACCCAGGGTTCATGATATAC 446
 Db 121 AGCATGGTATCCACTAGGAAGATAGTAGGTGCTACTGTTTACCCAGGGTTCATGATATAC 180

QY 447 CGTGCGCTTATTCATTTGATGATTTTAAATACATTAAGTAACTGTTTACATGAAGAGCT 506
 Db 181 CGTGCGCTTATTCATTTGATGATTTTAAATACATTAAGTAACTGTTTACATGAAGAGCT 240

QY 507 ATGTGTGTTCTTTCGACCACTTTTACCGGCTTACATCTATATCTACTTCTCTCTTAC 566
 Db 241 ATGTGTGTTCTTTCGACCACTTTTACCGGCTTACATCTATATCTACTTCTCTCTTAC 300

QY 567 AGAGAACTTTGGAACCAAGGAGCAGGACTTTTAGCTGCTTGTGTTTATTTGCTATTGTACC 626
 Db 301 AGAGAACTTTGGAACCAAGGAGCAGGACTTTTAGCTGCTTGTGTTTATTTGCTATTGTACC 360

QY 627 AGGCTACATATCTCGGTCTAGTAGTGGATCCCTTGTGATGAAGGATGCTATTGTTGC 686
 Db 361 AGGCTACATATCTCGGTCTAGTAGTGGATCCCTTGTGATGAAGGATGCTATTGTTGC 420

QY 687 ACTTCAGTTTACATATCTTTTATGGTAAATCTGTAAATCTGGTTCAGTTTGTGAC 746
 Db 421 ACTTCAGTTTACATATCTTTTATGGTAAATCTGTAAATCTGGTTCAGTTTGTGAC 480

QY 747 AATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGGGTGTTATGTTAT 806
 Db 481 AATGTGCTGCTGTTATCTTATTTCTATATGTTCTGCTGGGTGTTATGTTAT 540

QY 807 CATCAATCTTATTCCTACTGCATGTTATTTGTTGTTTACTGTATGAGAGATACAGAAAG 866
 Db 541 CATCAATCTTATTCCTACTGCATGTTATTTGTTGTTTACTGTATGAGAGATACAGAAAG 600

QY 867 AGTCTACATAGCATATGACATTTCTACATTTGGGTTTAAATTTATCAATGACATACC 926

Db 2281 ACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGCAGCCTGATTAACAGGAGACA 2340
 QY 2341 TTAGATACAAACCTCGAGTCACCAACATTTCCCAAAACAGAGAGTATTTCTCAAGAG 2400
 Db 2341 TTAGATACAAACCTCGAGTCACCAACATTTCCCAAAACAGAGAGTATTTCTCAAGAG 2400
 QY 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGGTTTTTAAGAAAGGCAAGAA 2460
 Db 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGGTTTTTAAGAAAGGCAAGAA 2460
 QY 2461 ATATCTAAGAGACTGTTTAA 2481
 Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 2
 US-09-945-527-62
 ; Sequence 62, Application US/09945527
 ; Publication No. US20030055588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: Nucleic Acid and Protein Homologs
 ; FILE REFERENCE: 35800/237985
 ; CURRENT APPLICATION NUMBER: US/09/945,527
 ; PRIORITY FILING DATE: 2001-08-29
 ; NUMBER OF SEQ ID NOS: 65
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 62
 ; LENGTH: 1828
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: 1767..1771
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-945-527-62

Query Match 43.1%; Score 1069; DB 10; Length 1828;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1069; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1413 GTCTGCAATGCTTTTCAATGTTTGGAGCACTATTGGGGGATGACATGAAAGGGA 1472
 Db 69 GTCTGCAATGCTTTTCAATGTTTGGAGCACTATTGGGGGATGACATGAAAGGGA 128
 QY 1473 AAATCCACCTGTGGAGGAGCAGCAGTGTAGGATGACAAAGAAACCAAGGAATTTGTA 1532
 Db 129 AAATCCACCTGTGGAGGAGCAGCAGTGTAGGATGACAAAGAAACCAAGGAATTTGTA 188
 QY 1533 TGATAGGAGGAGTAAAGTACAGAAACATCACTGAAAGGAAACCAAGAGGGATT 1592
 Db 189 TGATAGGAGGAGTAAAGTACAGAAACATCACTGAAAGGAAACCAAGAGGGATT 248
 QY 1593 AGSCCCTAATATAAAGCAATGTCACCATGTTGATGCTGATGATGATGATGATGCTGC 1652
 Db 249 AGSCCCTAATATAAAGCAATGTCACCATGTTGATGCTGATGATGATGATGATGCTGC 308
 QY 1653 TGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1712
 Db 309 TGTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 368
 QY 1713 ATACAAATCATGATGGCAGGAGAAATATCTTAGATGATTTTAGAAGCTTACTTTTGCT 1772
 Db 369 ATACAAATCATGATGGCAGGAGAAATATCTTAGATGATTTTAGAAGCTTACTTTTGCT 428
 QY 1773 AAGGCAAAATACAGATGACATGACAGTAAATGCTTGGTGGGATTTAGGCTATCAGAT 1832
 Db 429 AAGGCAAAATACAGATGACATGACAGTAAATGCTTGGTGGGATTTAGGCTATCAGAT 488
 QY 1833 AGCTGGAATGGCTAAATAGAACTACGTTGGTGGGATTAATACACCTGGAATAACAGCCACAT 1892

Db 489 AGCTGGAATGGCTAAATAGAACTACGTTGGTGGATTAATTAACACCTGGAATAACAGCCACAT 548
 QY 1893 AGCACTGGTGGGAAAGCTATGCTTTCTAATGAACAGCAGCCTATAAAATCATAGGAC 1952
 Db 549 AGCACTGGTGGGAAAGCTATGCTTTCTAATGAACAGCAGCCTATAAAATCATAGGAC 608
 QY 1953 TCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGATGA 2012
 Db 609 TCTAGATGTAGATTAATGTTTGGTATTTTGGAGGGGTTATTTGGCTATTTCTGGTGATGA 668
 QY 2013 TATCAACAAATTTCTCTGATGTTAGGATAGCTGAAGGAGAACATCCCAAGACATTCG 2072
 Db 669 TATCAACAAATTTCTCTGATGTTAGGATAGCTGAAGGAGAACATCCCAAGACATTCG 728
 QY 2073 GGAAGTGAATTTTACCCCAAGGAGAAATTCGTTAGCAAAAGCAGGATCCCTTAC 2132
 Db 729 GGAAGTGAATTTTACCCCAAGGAGAAATTCGTTAGCAAAAGCAGGATCCCTTAC 788
 QY 2133 TTTGTTGAATTTGCTTATGTATATAATGTCATCTACATGATTTGGAGAAATGCAGCTGGA 2192
 Db 789 TTTGTTGAATTTGCTTATGTATATAATGTCATCTACATGATTTGGAGAAATGCAGCTGGA 848
 QY 2193 TTTTCTCACACCCCAAGGTTTGAACGAAACAGTAATGCTGAGATTTGGAATTAAGACAT 2252
 Db 849 TTTTCTCACACCCCAAGGTTTGAACGAAACAGTAATGCTGAGATTTGGAATTAAGACAT 908
 QY 2253 TAAATTCAAACATTTTGGAGAGAGCCCTTTACATCAGAACACTGCTGTTGTTAGGATATAA 2312
 Db 909 TAAATTCAAACATTTTGGAGAGAGCCCTTTACATCAGAACACTGCTGTTGTTAGGATATAA 968
 QY 2313 AGTAAAGCAGCTGATTAACAGGAGAGACATTTAGATCACAACCTCGAGTCACCAACATTT 2372
 Db 969 AGTAAAGCAGCTGATTAACAGGAGAGACATTTAGATCACAACCTCGAGTCACCAACATTT 1028
 QY 2373 CCACAAACAGAGATTTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAATAA 2432
 Db 1029 CCACAAACAGAGATTTTGTCAAAGAGACTACCAAGAGAGCGTGGCTACATTAATAA 1088
 QY 2433 TAAGCTGGTTTTTAAAGAAAGCAAGAAATATCTTAAGAGAGCTGTTTAA 2481
 Db 1089 TAAGCTGGTTTTTAAAGAAAGCAAGAAATATCTTAAGAGAGCTGTTTAA 1137

RESULT 3
 US-10-106-698-330
 ; Sequence 330, Application US/10106698
 ; Publication No. US20030109690A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; PRIORITY FILING DATE: 2002-03-27
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 330
 ; LENGTH: 1209
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: (1203)..(1203)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-106-698-330

Query Match 42.5%; Score 1054; DB 15; Length 1209;
 Best Local Similarity 100.0%; Pred. No. 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 00:55:34 ; Search time 2195.32 Seconds
(without alignments)
6239.285 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgccccc.....tatctagaagactgtttaa 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4093002 seqs, 2760418825 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 2481 | 100.0 | 2481 | 15 | US-10-028-384-1 |
| 2 | 1089 | 43.1 | 1828 | 10 | US-09-945-527-62 |
| 3 | 1054 | 42.5 | 1209 | 15 | US-10-106-698-330 |
| 4 | 1054 | 42.5 | 1209 | 16 | US-10-264-237-412 |
| 5 | 1054 | 42.5 | 1543 | 10 | US-09-974-879-133 |
| 6 | 1054 | 42.5 | 1543 | 10 | US-09-305-736-133 |
| 7 | 1054 | 42.5 | 1543 | 10 | US-09-818-683-133 |
| 8 | 1054 | 42.5 | 1543 | 11 | US-09-818-683-133 |
| 9 | 1054 | 42.5 | 1543 | 16 | US-10-621-401-133 |
| 10 | 840 | 33.9 | 2660 | 16 | US-10-264-049-630 |
| 11 | 653 | 26.3 | 1114 | 16 | US-10-296-115-629 |
| 12 | 500 | 20.2 | 500 | 9 | US-09-998-598-1643 |

| | | | | | | | |
|----|------|------|------|-------------------|---------------------|--------------------|-------------------|
| 13 | 430 | 17.3 | 430 | 9 | US-09-878-178-717 | Sequence 717, App | |
| 14 | 430 | 17.3 | 430 | 13 | US-10-046-935-717 | Sequence 717, App | |
| 15 | 430 | 17.3 | 430 | 14 | US-10-146-503-717 | Sequence 717, App | |
| 16 | 402 | 16.2 | 558 | 14 | US-10-052-283-433 | Sequence 433, App | |
| 17 | 359 | 14.5 | 387 | 16 | US-10-276-774-173 | Sequence 173, App | |
| 18 | 351 | 14.1 | 616 | 9 | US-09-879-536-332 | Sequence 332, App | |
| 19 | 328 | 13.2 | 485 | 10 | US-09-918-995-11283 | Sequence 11283, A | |
| c | 20 | 12.7 | 743 | 10 | US-09-945-527-63 | Sequence 63, Appl | |
| 21 | 17.2 | 474 | 14 | US-10-052-283-454 | Sequence 454, App | | |
| 22 | 156 | 6.3 | 393 | 16 | US-10-276-774-323 | Sequence 323, App | |
| 23 | 135 | 5.4 | 312 | 9 | US-09-796-692-8870 | Sequence 8870, Ap | |
| 24 | 135 | 5.4 | 312 | 14 | US-10-040-862-8870 | Sequence 8870, Ap | |
| 25 | 135 | 5.4 | 312 | 16 | US-10-057-475B-8870 | Sequence 8870, Ap | |
| 26 | 135 | 5.4 | 312 | 16 | US-10-154-884B-8870 | Sequence 8870, Ap | |
| 27 | 135 | 5.4 | 312 | 17 | US-10-764-324-8870 | Sequence 8870, Ap | |
| c | 28 | 124 | 5.0 | 352 | 9 | US-09-878-178-1032 | Sequence 1032, Ap |
| 29 | 124 | 5.0 | 352 | 13 | US-10-046-935-1032 | Sequence 1032, Ap | |
| 30 | 124 | 5.0 | 352 | 14 | US-10-146-503-1032 | Sequence 1032, Ap | |
| 31 | 105 | 4.2 | 483 | 10 | US-09-918-995-23896 | Sequence 23896, A | |
| 32 | 65 | 2.6 | 2710 | 15 | US-10-028-384-3 | Sequence 3, Appli | |
| 33 | 44 | 1.8 | 637 | 15 | US-10-002-631C-183 | Sequence 183, App | |
| 34 | 44 | 1.8 | 640 | 15 | US-10-002-631C-179 | Sequence 179, App | |
| 35 | 27 | 1.1 | 424 | 10 | US-09-918-995-17769 | Sequence 17769, A | |
| c | 36 | 24 | 1.0 | 235 | 10 | US-09-945-527-61 | Sequence 61, Appl |
| 37 | 23 | 0.9 | 252 | 18 | US-10-128-558-54 | Sequence 54, Appl | |
| 38 | 23 | 0.9 | 307 | 9 | US-09-736-457-375 | Sequence 375, App | |
| 39 | 23 | 0.9 | 307 | 9 | US-09-736-457-588 | Sequence 588, App | |
| 40 | 23 | 0.9 | 307 | 9 | US-09-736-457-1271 | Sequence 1271, Ap | |
| 41 | 23 | 0.9 | 307 | 9 | US-09-902-941-375 | Sequence 375, App | |
| 42 | 23 | 0.9 | 307 | 9 | US-09-902-941-588 | Sequence 588, App | |
| 43 | 23 | 0.9 | 307 | 9 | US-09-902-941-1271 | Sequence 1271, Ap | |
| 44 | 23 | 0.9 | 307 | 9 | US-09-849-626-375 | Sequence 375, App | |
| 45 | 23 | 0.9 | 307 | 9 | US-09-849-626-588 | Sequence 588, App | |

ALIGNMENTS

RESULT 1
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

| | | | | | | | |
|-----------------------|---------|--|-------|------------|-----|--------|-------|
| Query Match | 100.0%; | Score | 2481; | DB | 15; | Length | 2481; |
| Best Local Similarity | 100.0%; | Pred. No. | 0; | | | | |
| Matches | 2481; | Conservative | 0; | Mismatches | 0; | Indels | 0; |
| Gaps | 0; | | | | | | |
| Qy | 1 | ATGGCGAGCCCTCGCCCGGAGCAAGTCTCCCTCAACTGTCCTCCGCGG | 60 | | | | |
| Db | 1 | ATGGCGAGCCCTCGCCCGGAGCAAGTCTCCCTCAACTGTCCTCCGCGG | 60 | | | | |
| Qy | 61 | AGTGGCTCATGCGCTTGGGAAACAGCGGCGCGGCGCCCGAGTGC | 120 | | | | |
| Db | 61 | AGTGGCTCATGCGCTTGGGAAACAGCGGCGCGGCGCCCGAGTGC | 120 | | | | |

~~DUPLICATE~~

Blank page

```
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 588
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-588

Query Match          0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
    |||||
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 13
US-09-614-124B-1271
; Sequence 1271, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-1271

Query Match          0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
    |||||
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 14
US-09-671-325-375
; Sequence 375, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 588
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-671-325-588

Query Match          0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
    |||||
Db 129 TGGTGGGATTATGGCTATCAGAT 151

Search completed: December 16, 2004, 07:21:29
Job time : 260.852 secs
```

```
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 375
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-671-325-375

Query Match          0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
    |||||
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 15
US-09-671-325-588
; Sequence 588, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 588
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-671-325-588

Query Match          0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
    |||||
Db 129 TGGTGGGATTATGGCTATCAGAT 151

Search completed: December 16, 2004, 07:21:29
Job time : 260.852 secs
```

APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedwick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 375
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-736-457-375

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 9

US-09-736-457-588
Sequence 588, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedwick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 588
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-736-457-588

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 10

US-09-736-457-1271
Sequence 1271, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong

APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedwick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1271
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-736-457-1271

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 11

US-09-614-124B-375
Sequence 375, Application US/09614124B
Patent No. 6630574
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedwick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.478C9
CURRENT APPLICATION NUMBER: US/09/614,124B
CURRENT FILING DATE: 2001-07-11
NUMBER OF SEQ ID NOS: 1668
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 375
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-614-124B-375

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1810 TGGTGGGATTATGGCTATCAGAT 1832
Db 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 12

US-09-614-124B-588
Sequence 588, Application US/09614124B
Patent No. 6630574
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.

PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 21090
LENGTH: 245
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-21090

Query Match 9.5%; Score 236; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.2e-112;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 CTACTTTCCCTGCTTACAAGAACTTTGGAACCAAGACGAGACTTTAGCTGCTGTT 610
DB 1 CTACTTTCCCTGCTTACAAGAACTTTGGAACCAAGACGAGACTTTAGCTGCTGTT 60

QY 611 TTATTGCTATTGTACAGGCTACATATCTGGTCACTAGTCTGGATCCCTTTGATAATGAAG 670
DB 61 TTATTGCTATTGTACAGGCTACATATCTGGTCACTAGTCTGGATCCCTTTGATAATGAAG 120

QY 671 GCATTGCTATTTTGCACCTTCAGTTCACATACACTATTATGGGTAAAAATCTGTAAAAAATCG 730
DB 121 GCATTGCTATTTTGCACCTTCAGTTCACATACACTATTATGGGTAAAAATCTGTAAAAAATCG 180

QY 731 GGTCAAGTTTTTGGACAATGTGCTGCTTATCTTATCTATGCTCTGCT 786
DB 181 GGTCAAGTTTTTGGACAATGTGCTGCTTATCTTATCTATGCTCTGCT 236

RESULT 5
US-09-702-705-375
Sequence 375, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 375
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-702-705-375

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 6
US-09-702-705-588
Sequence 588, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong

APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 588
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-702-705-588

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 7
US-09-702-705-1271
Sequence 1271, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1271
LENGTH: 307
TYPE: DNA
ORGANISM: Homo sapien
US-09-702-705-1271

Query Match 0.9%; Score 23; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TGGTGGGATTATGGCTATCAGAT 1832
DB 129 TGGTGGGATTATGGCTATCAGAT 151

RESULT 8
US-09-736-457-375
Sequence 375, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.

QY 1983 TGGAGGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGGATGTTAGGAT 2042
Db 301 TGGAGGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGGATGTTAGGAT 360
QY 2043 AGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCCAAGGGAGA 2102
Db 361 AGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGACTATTTTACCCCAAGGGAGA 420
QY 2103 ATTCCGTTAGACAAAGCAGATCCCTACTTTTGGTGAATTTGCTTATGTATTAATGTC 2162
Db 421 ATTCCGTTAGACAAAGCAGATCCCTACTTTTGGTGAATTTGCTTATGTATTAATGTC 480
QY 2163 ATACTACAGATTTGGAGAAATGCAGCT 2189
Db 481 ATACTACAGATTTGGAGAAATGCAGCT 507

RESULT 2

US-09-328-111-332

; Sequence 332, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/328,111

; EARLIER FILING DATE: 1999-06-08

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 332

; LENGTH: 616

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)...(616)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-332

Query Match 14.1%; Score 351; DB 3; Length 616;
Best Local Similarity 100.0%; Pred. No. 1.5e-172;
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GTACCAGGCTACATATCTCGGTACGTAGCTGATCCTTTGATATGAAGGCAATTCCTATT 691
Db 5 GTACCAGGCTACATATCTCGGTACGTAGCTGATCCTTTGATATGAAGGCAATTCCTATT 64
QY 682 TTGCACTTCAGTTCCATACATATTTATGGTAAATCTGTAAATCTGGGTCAAGTTT 741
Db 65 TTGCACTTCAGTTCCATACATATTTATGGTAAATCTGTAAATCTGGGTCAAGTTT 124
QY 742 TGGCAATGCTGCTGCTTATCTTATATGCTCTCTGCTGGGTGGTTATGTA 801
Db 125 TGGCAATGCTGCTGCTTATCTTATATGCTCTCTGCTGGGTGGTTATGTA 184
QY 802 TTTATCAATCTTATTCCTCACTGATGATTTTGGTTGTTACTGATCAGATACAGC 861
Db 185 TTTATCAATCTTATTCCTCACTGATGATTTTGGTTGTTACTGATCAGATACAGC 244

QY 862 AAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAAATATTATCAATGAG 921
Db 245 AAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAAATATTATCAATGAG 304
QY 922 ATACCTTTTGGGATTCAGCCCAATCAGACAAAGTGAACACATGGCAGCT 972
Db 305 ATACCTTTTGGGATTCAGCCCAATCAGACAAAGTGAACACATGGCAGCT 355

RESULT 3

US-09-513-999C-1438

; Sequence 1438, Application US/09513999C

; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm

; SEQ ID NO 1438

; LENGTH: 250

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 54..248

US-09-513-999C-1438

Query Match 10.1%; Score 250; DB 4; Length 250;
Best Local Similarity 100.0%; Pred. No. 6.2e-120;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 794 GTTATGTTATTTATCATCAATCTTATCCACTGATGTTTGTGTTTACTGATGAGA 853
Db 1 GTTATGTTATTTATCATCAATCTTATCCACTGATGTTTGTGTTTACTGATGAGA 60
QY 854 GATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAAATATT 913
Db 61 GATACAGCAAAAGAGTCTACATAGCATATAGCACTTTCTACATTTGGGTTTAAATATT 120
QY 914 CAATGAGATACCTTTTGTGGATTCCAGCCCAATCAGACAAAGTGAACACATGGCAGCTG 973
Db 121 CAATGAGATACCTTTTGTGGATTCCAGCCCAATCAGACAAAGTGAACACATGGCAGCTG 180
QY 974 CAGGTGTTTGTGATTTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAA 1033
Db 181 CAGGTGTTTGTGATTTGCTGCAAGCTTATGCTTTCTTTCAGTATCTGAGAGACCGATTAA 240
QY 1034 CAAAACAAGA 1043
Db 241 CAAAACAAGA 250

RESULT 4

US-09-513-999C-21090

; Sequence 21090, Application US/09513999C

; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:34:04 ; Search time 259.852 Seconds
(without alignments)
6786.440 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctggccccc.....tatctaagaagactgttttaa 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1 | 507 | 20.4 | 507 | US-09-513-999C-1965 | Sequence 1965, Ap |
| 2 | 351 | 14.1 | 616 | US-09-328-111-332 | Sequence 332, App |
| 3 | 250 | 10.1 | 250 | US-09-513-999C-1438 | Sequence 1438, App |
| 4 | 236 | 9.5 | 245 | US-09-513-999C-21090 | Sequence 21090, A |
| 5 | 23 | 0.9 | 307 | US-09-702-705-375 | Sequence 375, App |
| 6 | 23 | 0.9 | 307 | US-09-702-705-588 | Sequence 588, App |
| 7 | 23 | 0.9 | 307 | US-09-702-705-1271 | Sequence 1271, Ap |
| 8 | 23 | 0.9 | 307 | US-09-736-457-375 | Sequence 375, App |
| 9 | 23 | 0.9 | 307 | US-09-736-457-588 | Sequence 588, App |
| 10 | 23 | 0.9 | 307 | US-09-736-457-1271 | Sequence 1271, Ap |
| 11 | 23 | 0.9 | 307 | US-09-614-124B-375 | Sequence 375, App |
| 12 | 23 | 0.9 | 307 | US-09-614-124B-588 | Sequence 588, App |
| 13 | 23 | 0.9 | 307 | US-09-614-124B-1271 | Sequence 1271, Ap |
| 14 | 23 | 0.9 | 307 | US-09-671-325-375 | Sequence 375, App |
| 15 | 23 | 0.9 | 307 | US-09-671-325-588 | Sequence 588, App |
| 16 | 23 | 0.9 | 307 | US-09-671-325-1271 | Sequence 1271, Ap |
| 17 | 23 | 0.9 | 307 | US-09-589-184-375 | Sequence 375, App |
| 18 | 23 | 0.9 | 307 | US-09-589-184-588 | Sequence 588, App |
| 19 | 23 | 0.9 | 307 | US-09-589-184-1271 | Sequence 1271, Ap |
| 20 | 23 | 0.9 | 307 | US-09-658-824-375 | Sequence 375, App |
| 21 | 23 | 0.9 | 307 | US-09-658-824-588 | Sequence 588, App |
| 22 | 23 | 0.9 | 307 | US-09-658-824-1271 | Sequence 1271, Ap |
| 23 | 23 | 0.9 | 487 | US-09-702-705-1655 | Sequence 1655, Ap |
| 24 | 23 | 0.9 | 487 | US-09-736-457-1655 | Sequence 1655, Ap |
| 25 | 23 | 0.9 | 487 | US-09-614-124B-1655 | Sequence 1655, Ap |
| 26 | 23 | 0.9 | 487 | US-09-671-325-1655 | Sequence 1655, Ap |
| 27 | 21 | 0.8 | 480 | US-09-489-039A-5999 | Sequence 5999, Ap |

ALIGNMENTS

RESULT 1

US-09-513-999C-1965
Sequence 1965, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
PATENT NO. 6783961
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 1965
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: 122..505
US-09-513-999C-1965

| QY | 1683 | CTACTCTAGTCCAAAGTGTAGTCTGGCCTCATCAATCATGATGGCACCAGGATATCTT | 1742 |
|----|------|---|------|
| Db | 1 | CTACTCTAGTCCAAAGTGTAGTCTGGCCTCATCAATCATGATGGCACCAGGATATCTT | 60 |
| QY | 1743 | AGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAGAGT | 1802 |
| Db | 61 | AGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGCAGAGT | 120 |
| QY | 1803 | AATGCTCTGGTGGGATTTATGGCTATCAGATAGTGGGAATGGCTAATAGAACTACGTTGGT | 1862 |
| Db | 121 | AATGCTCTGGTGGGATTTATGGCTATCAGATAGTGGGAATGGCTAATAGAACTACGTTGGT | 180 |
| QY | 1863 | GGATAATAACACCTGGGATTAACAGCCACATAGACACTGGTGGGAAAAGCTATGCTCTTAA | 1922 |
| Db | 181 | GGATAATAACACCTGGGATTAACAGCCACATAGACACTGGTGGGAAAAGCTATGCTCTTAA | 240 |
| QY | 1923 | TGAACACAGCAGCTTATAAATCATGAGGACTCTAGATGTAGATTTGTTTGGTTATTTT | 1982 |
| Db | 241 | TGAACACAGCAGCTTATAAATCATGAGGACTCTAGATGTAGATTTGTTTGGTTATTTT | 300 |

Blank page
~~duplicate~~

| | | | |
|------|----|---|------|
| 581 | Db | TTGTGTAATTCGCTTATGTATAAAATGTCATACTACAGATTGGAGAAATGCGAGCTGGAT | 640 |
| 2194 | Qy | TTTCGTACACCCCCAGGTTTGTACCCGAACACAGTAATGCTGAGATTGGAAATTAAGGACATT | 2253 |
| 641 | Db | TTTCGTACACCCCCAGGTTTGTACCCGAACACAGTAATGCTGAGATTGGAAATTAAGGACATT | 700 |
| 2254 | Qy | AAATTTCAAAATTTGGAAGAAGCGCTTTTACATCAGAACCACTGCGCTGTTTGGATATATAAA | 2313 |
| 701 | Db | AAATTTCAAAATTTGGAAGAAGCGCTTTTACATCAGAACCACTGCGCTGTTTGGATATATAAA | 760 |
| 2314 | Qy | GTAAGCAGCTGATACACGGGAGACATTAGATACCAACCTCGAGTCACCAACATTTTC | 2373 |
| 761 | Db | GTAAGCAGCTGATACACGGGAGACATTAGATACCAACCTCGAGTCACCAACATTTTC | 820 |
| 2374 | Qy | CCAAAACAGAAGTATTGTGTCAAAAGAAGCTACAAAAAGGAAGCGTGGCTACATTAATAAAT | 2433 |
| 821 | Db | CCAAAACAGAAGTATTGTGTCAAAAGAAGCTACAAAAAGGAAGCGTGGCTACATTAATAAAT | 880 |
| 2434 | Qy | AAGCTGTTTTTAAAGAAAGCAAGAAAATATCTAAGAAGACTGTTTAA | 2481 |
| 881 | Db | AAGCTGTTTTTAAAGAAAGCAAGAAAATATCTAAGAAGACTGTTTAA | 928 |

Search completed: December 16, 2004, 00:55:22
Job time : 16666.4 secs

```

FEATURES
  source      Location/Qualifiers
    1. .2503
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"

ORIGIN
  Query Match      33.4%; Score 828; DB 6; Length 2503;
  Best Local Similarity 100.0%; Pred. No. 0;
  Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1654 GTCCACTGTACTCTGGGTCAACAGCAATGCCCTACTAGTCCAAAGTAGTCTCTGGCTCA 1713
  DB 101 GTCCACTGTACTCTGGGTCAACAGCAATGCCCTACTAGTCCAAAGTAGTCTCTGGCTCA 160
  QY 1714 TACATATCATGTGACACAGCAATATCTTAGATGATTTAGAGAGCTTACTTTGGCTA 1773
  DB 151 TACATATCATGTGACACAGCAATATCTTAGATGATTTAGAGAGCTTACTTTGGCTA 220
  QY 1774 AGGCAAAATACAGATGAACATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 1833
  DB 221 AGGCAAAATACAGATGAACATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 280
  QY 1834 GCTGAATGGCTAATAGCAATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 1893
  DB 281 GCTGAATGGCTAATAGCAATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 340
  QY 1894 GCCTGTGGGAAAGCTATGCTCTTAATGAACAGAGGCTTAAATATCAGAGGACT 1953
  DB 341 GCCTGTGGGAAAGCTATGCTCTTAATGAACAGAGGCTTAAATATCAGAGGACT 400
  QY 1954 CTAGATGTAGATTAATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGTATGAT 2013
  DB 401 CTAGATGTAGATTAATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGTATGAT 460
  QY 2014 ATCAACAAATTTCTCTGGATGTTAGATAGCTGAGAGAGACATCCAAAGACATTCGG 2073
  DB 461 ATCAACAAATTTCTCTGGATGTTAGATAGCTGAGAGAGACATCCAAAGACATTCGG 520
  QY 2074 GAAAGTGACTATTTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACT 2133
  DB 521 GAAAGTGACTATTTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACT 580
  QY 2134 TTGTTGAATGCTTATGTAATAATGCTATCATCAGATTTGGAGAAATGAGCTGGAT 2193
  DB 581 TTGTTGAATGCTTATGTAATAATGCTATCATCAGATTTGGAGAAATGAGCTGGAT 640
  QY 2194 TTTGCTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTTGGAATTAAGGACATT 2253
  DB 641 TTTGCTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTTGGAATTAAGGACATT 700
  QY 2254 AAATTCACAAATTTTGGAGAGGCTTTTACATCAGACACTGCTGTTAGGATATATAA 2313
  DB 701 AAATTCACAAATTTTGGAGAGGCTTTTACATCAGACACTGCTGTTAGGATATATAA 760
  QY 2314 GTAAAGCACCCTGTATACAGGAGACATTTAGATCAAAACCTCGAGTCAACCAATTTTC 2373
  DB 761 GTAAAGCACCCTGTATACAGGAGACATTTAGATCAAAACCTCGAGTCAACCAATTTTC 820
  QY 2374 CCAAAACAGAGATTTTGTCAAGAGACTTACCAAGAGAGGCTGCTTACATTAATAAT 2433
  DB 821 CCAAAACAGAGATTTTGTCAAGAGACTTACCAAGAGAGGCTGCTTACATTAATAAT 880
  QY 2434 AAGCTGGTTTTTAAAGAGGCAAGAAATATCTAAGAGACTGTTTAA 2481
  DB 881 AAGCTGGTTTTTAAAGAGGCAAGAAATATCTAAGAGACTGTTTAA 928

RESULT 15
BD137303
LOCUS      BD137303          2503 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION Human nucleic acid sequence originating in prostatic tissue.
ACCESSION BD137303

```

```

VERSION      BD137303.1      GI:23232248
KEYWORDS     JP 2002505878-A/64.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 2503)
AUTHORS     Specht,T., Hinzmann,B., Schmitt,A., Pilarsky,C., Dahl,E. and
             Rosenthal,A.
TITLE       Human nucleic acid sequence originating in prostatic tissue
JOURNAL     Patent: JP 2002505878-A 64 26-FEB-2002;
             METAGEN GESELLSCHAFT FUER GENOM FORSCHUNG MBH
COMMENT     OS Homo sapiens (human)
             PN JP 2002505878-A/64
             PD 26-FEB-2002
             PE 09-MAR-1999 JP 2000535742
             PF 10-MAR-1998 DE 198 11 194.0
             PI THOMAS SPECHT, BERND HINZMANN, ARMIN SCHMITT, CHRISTIAN PILARSKY,
             PI EDGAR DAHL,
             PI ANDRE ROSENTHAL
             PC C12N15/09, A61K38/00, A61K39/395, A61K48/00, A61P13/08,
             PC A61P35/00,
             PC C07K14/47, C07K16/18, C12N1/19, C12N1/21, C12N5/10, C12N15/02, PC
             C12P21/02.
             PC C12P21/08, G01N33/53, C12N15/00, A61K37/02, C12N5/00, C12N15/00 CC
             Human nucleic acid sequence originating in prostatic tissue PH
Key          Location/Qualifiers
FT          source      1. .2503
             /organism="Homo sapiens (human)".
FEATURES
  source      Location/Qualifiers
    1. .2503
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"

ORIGIN
  Query Match      33.4%; Score 828; DB 6; Length 2503;
  Best Local Similarity 100.0%; Pred. No. 0;
  Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1654 GTCCACTGTACTCTGGGTCAACAGCAATGCCCTACTAGTCCAAAGTAGTCTCTGGCTCA 1713
  DB 101 GTCCACTGTACTCTGGGTCAACAGCAATGCCCTACTAGTCCAAAGTAGTCTCTGGCTCA 160
  QY 1714 TACATATCATGTGACACAGCAATATCTTAGATGATTTAGAGAGCTTACTTTGGCTA 1773
  DB 161 TACATATCATGTGACACAGCAATATCTTAGATGATTTAGAGAGCTTACTTTGGCTA 220
  QY 1774 AGGCAAAATACAGATGAACATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 1833
  DB 221 AGGCAAAATACAGATGAACATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 280
  QY 1834 GCTGAATGGCTAATAGCAATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 1893
  DB 281 GCTGAATGGCTAATAGCAATGACAGCAATATCTTGTGGATTTATGGCTATCAGATA 340
  QY 1894 GCCTGTGGGAAAGCTATGCTCTTAATGAACAGAGGCTTAAATATCAGAGGACT 1953
  DB 341 GCCTGTGGGAAAGCTATGCTCTTAATGAACAGAGGCTTAAATATCAGAGGACT 400
  QY 1954 CTAGATGTAGATTAATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGTATGAT 2013
  DB 401 CTAGATGTAGATTAATGTTTGGTATTTTGGAGGGTATTTGGCTATTTCTGGTATGAT 460
  QY 2014 ATCAACAAATTTCTCTGGATGTTAGATAGCTGAGAGAGACATCCAAAGACATTCGG 2073
  DB 461 ATCAACAAATTTCTCTGGATGTTAGATAGCTGAGAGAGACATCCAAAGACATTCGG 520
  QY 2074 GAAAGTGACTATTTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACT 2133
  DB 521 GAAAGTGACTATTTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGGATCCCTACT 580
  QY 2134 TTGTTGAATGCTTATGTAATAATGCTATCATCAGATTTGGAGAAATGAGCTGGAT 2193
  DB 581 TTGTTGAATGCTTATGTAATAATGCTATCATCAGATTTGGAGAAATGAGCTGGAT 640
  QY 2194 TTTGCTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTTGGAATTAAGGACATT 2253
  DB 641 TTTGCTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTTGGAATTAAGGACATT 700
  QY 2254 AAATTCACAAATTTTGGAGAGGCTTTTACATCAGACACTGCTGTTAGGATATATAA 2313
  DB 701 AAATTCACAAATTTTGGAGAGGCTTTTACATCAGACACTGCTGTTAGGATATATAA 760
  QY 2314 GTAAAGCACCCTGTATACAGGAGACATTTAGATCAAAACCTCGAGTCAACCAATTTTC 2373
  DB 761 GTAAAGCACCCTGTATACAGGAGACATTTAGATCAAAACCTCGAGTCAACCAATTTTC 820
  QY 2374 CCAAAACAGAGATTTTGTCAAGAGACTTACCAAGAGAGGCTGCTTACATTAATAAT 2433
  DB 821 CCAAAACAGAGATTTTGTCAAGAGACTTACCAAGAGAGGCTGCTTACATTAATAAT 880
  QY 2434 AAGCTGGTTTTTAAAGAGGCAAGAAATATCTAAGAGACTGTTTAA 2481
  DB 881 AAGCTGGTTTTTAAAGAGGCAAGAAATATCTAAGAGACTGTTTAA 928

```

Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Mori, S., Moni, Y., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shinizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Suiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumei, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, H., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Sato, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T., and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
14702039

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Nabekura, T., Ishii, S., Kawai, Y., Saito, K., Yamamoto, G., Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y., and Oshima, A. NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2284)
Isogai, T. and Otsuki, T.
Direct Submission

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

Submitted (10-MAY-2003) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yata, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES
source

Location/Qualifiers
1..2284
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1003596"
/tissue_type="placenta"
/clone_lib="PLACE1"
/note="Cloning vector: pME18SFL3"

CDS

3..863
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:14042728"
/translation="MLMLMLMFAVHCTMTNSAVSPSVVLSYNHDEGRNLDPE
REAYFMRONTDSHPVMSWDYGYOIAQNRFTLVNNTWNSHIALVCKMSSNE
TAAYKIMRLDVLVIFGVISGDDINKPLWNRVAGEHPDRRESDFTPQS
EPRVKAQSTPLNCLMKNSYRFGEMQLDPTPPGDFRNBAGNKDKFKHLEE
APFSEWLVRIYKAPDNRETLNLDHKEPVTNIPFKQVLSKTKTKRKGKYLKKNLVEK
KGKISKIKTV"

ORIGIN

Query Match 34.7%; Score 861; DB 9; Length 2284;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1621 ATGTGATGCTGATGCTATTGATGATGTTTGTCTGCTCACTGCTGCTGCTCAAGCAAT 1680
DB 3 ATGTGATGCTGATGCTATTGATGATGTTTGTCTGCTCACTGCTGCTGCTCAAGCAAT 52
QY 1681 GCCTACTCTAGTCCCAAGTGTAGTCCCTGCTCATCAATCATGATGGCCACAGGAATATC 1740
DB 63 GCCTACTCTAGTCCCAAGTGTAGTCCCTGCTCATCAATCATGATGGCCACAGGAATATC 122

QY 1741 TTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGCGAAAATACAGATGAACATGCACGA 1800
DB 123 TTAGATGATTTTAGAAGAGCTTACTTTTGGCTAAGCGAAAATACAGATGAACATGCACGA 182
QY 1801 GTAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTTG 1860
DB 183 GTAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGTTG 242
QY 1861 GTGGATAATAACACCTGGAATAACAGCCACATAGCACTGTGTGGGAAAAGCTATGCTTCT 1920
DB 243 GTGGATAATAACACCTGGAATAACAGCCACATAGCACTGTGTGGGAAAAGCTATGCTTCT 302
QY 1921 ATGAAAAGCAGCAGCTATAAATCATCAGCACTCTAGATAGCTATGTTTGGTATT 1980
DB 303 AATGAAAAGCAGCAGCTATAAATCATCAGCACTCTAGATAGCTATGTTTGGTATT 362
QY 1981 TTTGGAGGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAG 2040
DB 363 TTTGGAGGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGGATGGTTAG 422
QY 2041 ATAGCTGAAGGAGACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGA 2100
DB 423 ATAGCTGAAGGAGACATCCCAAGACATTCGGGAAAGTGAATTTTACCCACAGGGA 482
QY 2101 GAATTCCTGTAGCAAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGATATAAATG 2160
DB 483 GAATTCCTGTAGCAAAAGCAGGATCCCTACTTTTGTGAATTCCTTATGATATAAATG 542
QY 2161 TCATATCACTAGTTTGGGAAATCGAGTCTGATATTCGTACACCCACAGGTTTTCACCGA 2220
DB 543 TCATATCACTAGTTTGGGAAATCGAGTCTGATATTCGTACACCCACAGGTTTTCACCGA 602
QY 2221 ACACCTAATGCTGAGATTGGAATTAAGGACATTAATTAACAACTTGAAGAAGCTTT 2280
DB 603 ACACCTAATGCTGAGATTGGAATTAAGGACATTAATTAACAACTTGAAGAAGCTTT 662
QY 2281 ACATCAGACACTGGCTGTGTAGGATATATAAGTAAAGCACTGATACAGGGAGACA 2340
DB 663 ACATCAGACACTGGCTGTGTAGGATATATAAGTAAAGCACTGATACAGGGAGACA 722
QY 2341 TTAGATCAAAACCTCGAGTCACCAACATTTCCCAAAACAGAAAGTATTGTCAAAAGAG 2400
DB 723 TTAGATCAAAACCTCGAGTCACCAACATTTCCCAAAACAGAAAGTATTGTCAAAAGAG 782
QY 2401 ACTACCAAAAGGAGCGTGGCTACATATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 2460
DB 783 ACTACCAAAAGGAGCGTGGCTACATATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 842
QY 2461 ATATCTAAGAGAGCTGTTTAA 2481
DB 843 ATATCTAAGAGAGCTGTTTAA 863

RESULT 14

AX017997
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

AX017997
Sequence 232 from Patent WO9946375.
AX017997
AX017997.1 GI:10042496
Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

QY 1621

DB 3

QY 1681

DB 63

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Schmitt, A., Specht, T., Dahl, E., Hinzmann, B., Rosenthal, A. and
Pilarczyk, C.
Human nucleic acid sequences from prostate tissue
Patent: WO 9946375-A 232 16-SEP-1999;
SCHMITT THOMAS (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)

```

Db      663 ACATCAGAACACTGGCTGTTAGGATATATAAAGTAAAGACACCTGATTAACAGGGAGACA 722
Qy      2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAG 2400
Db      723 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAG 782
Qy      2401 ACTACCAAAAGGAGCGTGGCTAGCATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 2460
Db      783 ACTACCAAAAGGAGCGTGGCTAGCATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 842
Qy      2461 ATATCTAAGAGACTGTTTAA 2481
Db      843 ATATCTAAGAGACTGTTTAA 863

RESULT 12
BD160013
LOCUS      2284 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD160013
VERSION    BD160013.1 GI:27865771
KEYWORDS   JP 2002191363-A/14856.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2284)
Ota,T., Isegai,T., Nishikawa,T., Hayaishi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 14856 09-JUL-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/14856
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOgai,TEISUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TEISUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12NS/ PC
10,
PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS
Location/Qualifiers
1..2284
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
Source
Query Match 34.7%; Score 861; DB 6; Length 2284;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 861; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
Qy      1621 ATGTTGATGCTGATGCTATGATGATCTTTGCTGTCACCTGACCTGGTCTACAGCAAT 1680
Db      3 ATGTTGATGCTGATGCTATGATGATCTTTGCTGTCACCTGACCTGGTCTACAGCAAT 62
Qy      1681 GCCTACTCTAGTCCAAAGTGTAGTCTCTGGCTCATACATCATGATGGCACCAGGAATATC 1740
Db      63 GCCTACTCTAGTCCAAAGTGTAGTCTCTGGCTCATACATCATGATGGCACCAGGAATATC 122
Qy      1741 TTAGATGATTTTATAGAGAGCTTACTTTTGGCTAAGCCAAATACAGATCAATCAGCA 1800
Db      123 TTAGATGATTTTATAGAGAGCTTACTTTTGGCTAAGCCAAATACAGATCAATCAGCA 182
Qy      1801 GTAAATGCTTGGTGGGATTAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 1860
Db      183 GTAAATGCTTGGTGGGATTAATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 242

```

```

Qy      1861 GTGGATATATACACCTCGGATTAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTTCT 1920
Db      243 GTGGATATATACACCTCGGATTAACAGCCACATAGCACTGGTGGGAAAAGCTATGCTTCT 302
Qy      1921 ATGAAAACAGCAGCCTTATAAATCATAGAGGACTCTAGATGTAGATTATGTTGGTTAT 1980
Db      303 AATGAAAACAGCAGCCTTATAAATCATAGAGGACTCTAGATGTAGATTATGTTGGTTAT 362
Qy      1981 TTTGAGGGGGTTATTTGGCTATTTCTGGTGATGATCAACAAATTTCTCTGGATGGTTAGG 2040
Db      363 TTTGAGGGGGTTATTTGGCTATTTCTGGTGATGATCAACAAATTTCTCTGGATGGTTAGG 422
Qy      2041 ATAGCTGAAGGAGAAATCCCAAGACATTCGGAAAGTGACTATTTTACCCACAGGGA 2100
Db      423 ATAGCTGAAGGAGAAATCCCAAGACATTCGGAAAGTGACTATTTTACCCACAGGGA 482
Qy      2101 GAATTCCTGGTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTCCTTATGTATAAAATG 2160
Db      483 GAATTCCTGGTGTAGACAAAGCAGGATCCCTACTTTTGTGTAATTCCTTATGTATAAAATG 542
Qy      2161 TCATCTACTCAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCACAGGTTTGGACCGA 2220
Db      543 TCATCTACTCAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCACAGGTTTGGACCGA 602
Qy      2221 ACACTGTAATGCTGAGATTTGGAATTAAGGACATTTAAATTCAAACATTTTGGAAAGCCTTT 2280
Db      603 ACACTGTAATGCTGAGATTTGGAATTAAGGACATTTAAATTCAAACATTTTGGAAAGCCTTT 662
Qy      2281 ACATCAGAACACTGGCTGTTAGGATATATAAAGCTAAAGCAGCTGATTAACAGGGAGACA 2340
Db      663 ACATCAGAACACTGGCTGTTAGGATATATAAAGCTAAAGCAGCTGATTAACAGGGAGACA 722
Qy      2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAGAG 2400
Db      723 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGATTTTGTCAAGAGAG 782
Qy      2401 ACTACCAAAAGGAGCGTGGCTGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 2460
Db      783 ACTACCAAAAGGAGCGTGGCTGCTACATTAATAAATAAGCTGTTTAAAGAAAGGCAAGAAA 842
Qy      2461 ATATCTAAGAGACTGTTTAA 2481
Db      843 ATATCTAAGAGACTGTTTAA 863

RESULT 13
AK027789
LOCUS      2284 bp mRNA linear PRI 30-JAN-2004
DEFINITION Homo sapiens cDNA FLJ14883 fis, clone PLACE1003596, moderately
similar to OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT.
ACCESSION AK027789
VERSION    AK027789.1 GI:14042727
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Oca,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayaishi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahara,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M.,
Shitatori,A., Sudo,H., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikawa,E., Onura,Y., Abe,K., Kamihara,K., Katsuma,N., Sato,K.,
Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hirooka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yoshida,M., Houta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arifa,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,

```


[illegible]

```
PI KOJI HAYASHI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68, C12P21/08, C12N15/00, C12N5/00 CC
Secretory protein or membrane protein
FH Key Location/Qualifiers
FT CDS (94), (951).
FEATURES
    source
        1..2510
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
ORIGIN
    Query Match 38.5%; Score 954; DB 6; Length 2510;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1528 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACAACTGAAGAG 1587
DB 1 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACAACTGAAGAG 60
QY 1588 GGATTAGCCCTAATATAAAAGCATTTGCCACATGTCATGCTGATGCTATTGATGATG 1647
DB 61 GGATTAGCCCTAATATAAAAGCATTTGCCACATGTCATGCTGATGCTATTGATGATG 120
QY 1648 TTTGCTGTCCACTGCTACTCGGTGACAAAGCAATGCTACTAGTCCAAAGTGTAGCTGTG 1707
DB 121 TTTGCTGTCCACTGCTACTCGGTGACAAAGCAATGCTACTAGTCCAAAGTGTAGCTGTG 180
QY 1708 GCCTCATACATCATGATGTCACAGCAATATCTTAGATGATTTTAGAGAGCTTACTTTT 1767
DB 181 GCCTCATACATCATGATGTCACAGCAATATCTTAGATGATTTTAGAGAGCTTACTTTT 240
QY 1768 TGGCTAAGGCAAAATACAGATGAACATGACAGATTAATCTTGTGGGATTTATGCGTAT 1827
DB 241 TGGCTAAGGCAAAATACAGATGAACATGACAGATTAATCTTGTGGGATTTATGCGTAT 300
QY 1828 CAGATAGCTGAATGGCTAATAGAACTACCTGCTGGTGAATAAACAACCTGGAAATACAGC 1887
DB 301 CAGATAGCTGAATGGCTAATAGAACTACCTGCTGGTGAATAAACAACCTGGAAATACAGC 360
QY 1888 CACATAGACTGTGGGAAAGCTATGCTCTTCTTAATGAACACAGAGCTTATAAATCATG 1947
DB 361 CACATAGACTGTGGGAAAGCTATGCTCTTCTTAATGAACACAGAGCTTATAAATCATG 420
QY 1948 AGGACTCTAGATGATGATTAATGTTTGGTGTATTTTGGAGGGGTATTGGCTATTCTGCT 2007
DB 421 AGGACTCTAGATGATGATTAATGTTTGGTGTATTTTGGAGGGGTATTGGCTATTCTGCT 480
QY 2008 GATGATATCAACAAATTTCTCTGGATGTTAGATAGCTGAAGGAGAAACATCCCAAGAC 2067
DB 481 GATGATATCAACAAATTTCTCTGGATGTTAGATAGCTGAAGGAGAAACATCCCAAGAC 540
QY 2068 ATTCCGGAAGTGAATATTTTACCCACAGGGAGATTCCTGTTAGACAAAGCAGATCC 2127
DB 541 ATTCCGGAAGTGAATATTTTACCCACAGGGAGATTCCTGTTAGACAAAGCAGATCC 600
QY 2128 CCTACTTTGTTGAATTCCTTATGTTAATAAATGCTATACAGATTTGGAGAAATGAG 2187
DB 601 CCTACTTTGTTGAATTCCTTATGTTAATAAATGCTATACAGATTTGGAGAAATGAG 660
QY 2188 CTGGATTTTCGTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTCGAAATAG 2247
DB 661 CTGGATTTTCGTACACCCCGAGTTTTCACCGAACACGTAATGCTGAGATTCGAAATAG 720
QY 2248 GACATTAATTAACATTTGGAGAGAGCTTTTACATCAGACATCTGCTGTTTAGGATA 2307
DB 721 GACATTAATTAACATTTGGAGAGAGCTTTTACATCAGACATCTGCTGTTTAGGATA 780
QY 2308 TATAAGTAAAGACCTGATTAACAGGGAGACATTAGATCAACAACCTGCTGATCCACCAAC 2367
DB 1 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACAACTGAAGAG 60

Db 781 TATAAGTAAAGACCTGATTAACAGGGAGACATTAGATCAACAACCTGCTGATCCACCAAC 840
QY 2368 ATTTTCCCAAAACAGAAAGTATTGTCAAAAGAAAGACTACCAAAAGAAAGCGTGGTACATT 2427
DB 841 ATTTTCCCAAAACAGAAAGTATTGTCAAAAGAAAGACTACCAAAAGAAAGCGTGGTACATT 900
QY 2428 AAAAAATAAGCTGTTTAAAGAAAGCGAAGAAATATCTTAAGAAAGACTGTTTAA 2481
DB 901 AAAAAATAAGCTGTTTAAAGAAAGCGAAGAAATATCTTAAGAAAGACTGTTTAA 954

RESULT 10
AK075380 2510 bp mRNA linear PRI 03-SEP-2002
LOCUS Homo sapiens cDNA PSEC0070 fis, clone NT2RP2001508, moderately
DEFINITION similar to OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT.
ACCESSION AK075380.1 GI:22761428
VERSION AK075380.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Kawai-Hio, Y., Hayashi, K.,
Ishii, S., Saiko, K., Yamamoto, J., Wakamatsu, A., Nagai, T.,
Nakamura, Y., Nagahara, K., Sugano, S. and Isogai, T.
TITLE HRI human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2510)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
[E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986]
HRI human cDNA sequencing project; cDNA 5' - 3' -end one pass
sequencing, clone selection and full insert sequencing; Helix
Research Institute (supported by Japan Key Technology Center etc.);
cDNA library construction; Institute of Medical Science, University
of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES
    source
        1..2510
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /c1one="NT2RP2001508"
            /cell_line="NT2"
            /cell_type="teratocarcinoma"
            /clone_lib="NT2R22"
            /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
            precursor cells after 2-weeks retinoic acid (RA)
            induction"
            94..954
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="BAC11581.1"
            /db_xref="GI:22761428"
            /translation="WLMMLNMFVHCTWTWTSNAYSPSVVLASVYHDSNRLILDF
            RYAYFWLQNDIEHARVMSWMDYQYAGMANRTLLVDNNTWNNSHIALVKAMSSNE
            TAAKIMETLDDVYLVIFGVIGVGGDINKFLMWVIAEGEPKDIRSDYTPQG
            EFRVKGAGSLTNCLMKNKYRFGMDLFTPTGDFRTRNAEIGNKDKFKHLEE
            APTSHVLVRIYKVKAPDNRTLDHPRVTNIFPKYKLSKTKTKRGRYIKNKLVEK
            KGKTSKXTV"
            CDS
            ORIGIN
            Query Match 38.5%; Score 954; DB 9; Length 2510;
            Best Local Similarity 100.0%; Pred. No. 0;
            Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1528 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACAACTGAAGAG 1587
DB 1 TTGTATGATAGGCGAGTAAGTGAGGAAACATGCAACATGAAACAGGAAACAACTGAAGAG 60
```

Db 721 GCTGAGATTGGAATTAAGGACATTAATTCACAACTTTGGAGAGAGCCCTTTACATCAGAA 780

Qy 2290 CACTGGCTTTAGGATATATAAGTAAAGCACTGATAACAGGGAGACATTTAGATCAC 2349

Db 781 CACTGGCTTTAGGATATATAAGTAAAGCACTGATAACAGGGAGACATTTAGATCAC 840

Qy 2350 AAACCTCGAGTACCAACATTTTCCCAAAACAGAGTATTTCTCAAGAGAGACATCCAAA 2409

Db 841 AAACCTCGAGTACCAACATTTTCCCAAAACAGAGTATTTCTCAAGAGAGACATCCAAA 900

Qy 2410 AGGAAGCGTGGCTACATTAATAAATAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAG 2469

Db 901 AGGAAGCGTGGCTACATTAATAAATAGCTGGTTTTTAAGAAAGGCAAGAAATATCTAAG 960

Qy 2470 AAGACTGTTTAA 2481

Db 961 AAGACTGTTTAA 972

RESULT 8

AX136135

LOCUS 2510 bp DNA linear PAT 30-MAY-2001

DEFINITION Sequence 57 from Patent EP1067182.

ACCESSION AX136135

VERSION AX136135.1 GI:14272543

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and Hayashi, K.

TITLE Secretory protein or membrane protein

JOURNAL Patent: EP 1067182-A 57 10-JAN-2001; Helix Research Institute (JP)

FEATURES

source 1..2510

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

94..954

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC39710.1"

/db_xref="GI:1427254"

/translation="MIMLLMFMFAVHCTWTSNAYSSPSVVLASYNHGDPNILDPE REAFVLRONTDEHAFVMSWDYGYQIAGMANRTLVNDNTMNSHIALVKKMSSNE TAAKIMRLDLDVYLVIEGGVIGSGDDINKFLMWVIAEGEHPKIRSDYFTPG EFRVDRAGSFTLNCMLYKMSYRFGEMQLDFTPPDFTRNNAIGNKDIKPKHLE AFTSEHLVRIYKVPADPNRETLDHKPRVTNIFPKQKYLKSKTKTKRKGVIKDKLVFK KGKKISKKTIV"

CDS

Query Match 38.5%; Score 954; DB 6; Length 2510;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1528 TTGTATGATAGGACAGTAAAGTGAAGAAACATGCACTGAACAGGAAAAAAGTGAAG 1587

Db 1 TTGTATGATAGGACAGTAAAGTGAAGAAACATGCACTGAACAGGAAAAAAGTGAAG 60

Qy 1588 GGATTAGGCGCTTAATATAAAGCATTTGACCATGTTGATGCTCTGATGATGATG 1647

Db 61 GGATTAGGCGCTTAATATAAAGCATTTGACCATGTTGATGCTCTGATGATGATG 120

Qy 1648 TTGTCTGTCACATGTTACCTGGGTGCAACAAGATGCTCTAGTCCAAAGTGTAGTCTG 1707

Db 121 TTGTCTGTCACATGTTACCTGGGTGCAACAAGATGCTCTAGTCCAAAGTGTAGTCTG 180

Qy 1708 GCCTCATACATCATGATGTCACAGCAATATCTTAGATGATTTTAGAGAGAGCTTACTTT 1767

Db 181 GCCTCATACATCATGATGTCACAGCAATATCTTAGATGATTTTAGAGAGAGCTTACTTT 240

Qy 1769 TGGCTAAGGCAAAATACAGATGAACATGCACAGAGTAATGTCTTGGTGGGATTTAGGCTAT 1827

Db 241 TGGCTAAGGCAAAATACAGATGAACATGCACAGAGTAATGTCTTGGTGGGATTTAGGCTAT 300

Qy 1828 CAGATAGCTGGAAATGGCTTAATAGAACTACGTTGGTGGTAAATAACACCTCGAAATAACAGC 1887

Db 301 CAGATAGCTGGAAATGGCTTAATAGAACTACGTTGGTGGTAAATAACACCTCGAAATAACAGC 360

Qy 1889 CACATAGCACTGGTGGGAAAGCTATGTCTTCTAATGAAACAGCAGCCTTATAAATCATG 1947

Db 361 CACATAGCACTGGTGGGAAAGCTATGTCTTCTAATGAAACAGCAGCCTTATAAATCATG 420

Qy 1948 AGGACTCTAGATGCTAGATTAATGTTTGGTATTTTTCGAGGGGTATTGCTGCTATTCTGGT 2007

Db 421 AGGACTCTAGATGCTAGATTAATGTTTGGTATTTTTCGAGGGGTATTGCTGCTATTCTGGT 480

Qy 2008 GATGATATCAACAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGACATCCCAAGAC 2067

Db 481 GATGATATCAACAATTTCTCTGGATGGTTAGGATAGCTGAAGGAGACATCCCAAGAC 540

Qy 2068 ATTCCGGAAAGTGACTATTTTACCCCAAGGAGAAATTCCTGTGTAGACAAAGCAGGATCC 2127

Db 541 ATTCCGGAAAGTGACTATTTTACCCCAAGGAGAAATTCCTGTGTAGACAAAGCAGGATCC 600

Qy 2128 CCTACTTTTGGTAATGGCTTATGTATAAATGTCTACTACTACAGATTTGGAGAAATGCAG 2187

Db 601 CCTACTTTTGGTAATGGCTTATGTATAAATGTCTACTACTACAGATTTGGAGAAATGCAG 660

Qy 2188 CTGGATTTTCTGTACACCCCAAGGTTTTGACCCCAAGGAGAAATTCCTGTGTAGACAAAGCAGGATCC 2247

Db 661 CTGGATTTTCTGTACACCCCAAGGTTTTGACCCCAAGGAGAAATTCCTGTGTAGACAAAGCAGGATCC 720

Qy 2248 GACATTAATTAACAACATTTGGAAGAGCCCTTTTACATCAGAAACACTGGCTTTGTAGATA 2307

Db 721 GACATTAATTAACAACATTTGGAAGAGCCCTTTTACATCAGAAACACTGGCTTTGTAGATA 780

Qy 2308 TATAAAGTAAAGACACCTGTATAACAGGAGACATTTAGATCACAACCTCCAGTCAACCAAC 2367

Db 781 TATAAAGTAAAGACACCTGTATAACAGGAGACATTTAGATCACAACCTCCAGTCAACCAAC 840

Qy 2368 ATTTTCCCAAAACAGAGTATTTTGTCAAGAGACTACCAAGAGAGCGTGGCTACATT 2427

Db 841 ATTTTCCCAAAACAGAGTATTTTGTCAAGAGACTACCAAGAGAGCGTGGCTACATT 900

Qy 2428 AAAAAATAGCTGGTTTTTAAAGAGGCAAGAAAAATATCTAAGAGAGCTGTTTAA 2481

Db 901 AAAAAATAGCTGGTTTTTAAAGAGGCAAGAAAAATATCTAAGAGAGCTGTTTAA 954

RESULT 9

BD123520

LOCUS 2510 bp DNA linear PAT 18-SEP-2002

DEFINITION Secretory protein or membrane protein.

ACCESSION BD123520

VERSION BD123520.1 GI:23218465

KEYWORDS JP 2002017376-A/29.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2510)

AUTHORS Ota, T., Isogai, T., Nishikawa, T., Kawai, Y., Sugiyama, T. and Hayashi, K.

TITLE Secretory protein or membrane protein

JOURNAL Patent: JP 2002017376-A 29 22-JAN-2002; HELIX RESEARCH INSTITUTE

COMMENT OS Homo sapiens (human)

PN JP 2002017376-A/29

PD 22-JAN-2002

PF 07-JUL-2000 JP 2000253173

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU SUGIYAMA,

QY 1690 AGTCCAAAGTGATGCTGGCTCATACATCATGATGGACACAGGAATCTTAGATGAT 1749
 Db 181 AGTCCAAAGTGATGCTGGCTCATACATCATGATGGACACAGGAATCTTAGATGAT 240
 QY 1750 TTTAGAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGTCT 1809
 Db 241 TTTAGAGAGGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGTCT 300
 QY 1810 TGGTGGGATATGCTATCAGATAGCTGGATGGCTAATAGAACTACGTTGGTGAATAT 1869
 Db 301 TGGTGGGATATGCTATCAGATAGCTGGATGGCTAATAGAACTACGTTGGTGAATAT 360
 QY 1870 AACACCTGGATTAACACCCATAGCATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 1929
 Db 361 AACACCTGGATTAACACCCATAGCATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 420
 QY 1930 GCAGCTTAATAATCATGAGACTCTAGATGATAGATTAATGTTGGTATATTTTGGAGGG 1989
 Db 421 GCAGCTTAATAATCATGAGACTCTAGATGATAGATTAATGTTGGTATATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 2049
 Db 481 GTTATTGGCTATCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 2109
 Db 541 GGAGACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 600
 QY 2110 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATTTGATTAATAAATGTCTACTATAC 2169
 Db 601 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATTTGATTAATAAATGTCTACTATAC 660
 QY 2170 AGATTGGAGAAATGAGCTGGATTTTCTGACACCCAGGTTTGGACGACAGTAAT 2229
 Db 661 AGATTGGAGAAATGAGCTGGATTTTCTGACACCCAGGTTTGGACGACAGTAAT 720
 QY 2230 GCTGAGATTGGAAATTAAGACATTAATAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 2289
 Db 721 GCTGAGATTGGAAATTAAGACATTAATAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 780
 QY 2290 CACTGGCTCTTAGGATATTAAGTAAGACCTGATAACAGGAGACATTAGATCAC 2349
 Db 781 CACTGGCTCTTAGGATATTAAGTAAGACCTGATAACAGGAGACATTAGATCAC 840
 QY 2350 AAACCTCGAGTACCAACATTTTCCCAAAACAGAGTATTTGTCAAAGAGACTACCAA 2409
 Db 841 AAACCTCGAGTACCAACATTTTCCCAAAACAGAGTATTTGTCAAAGAGACTACCAA 900
 QY 2410 AGGAAGCTGGCTACATTAATAATAGCTGGTTTAAAGAGGCAAGAAATATCTAAG 2469
 Db 901 AGGAAGCTGGCTACATTAATAATAGCTGGTTTAAAGAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTCTTTAA 2481
 Db 961 AAGACTCTTTAA 972

RESULT 7
 LOCUS B063986 2546 bp DNA linear PAT 27-AUG-2002
 DEFINITION Secreted proteins and polynucleotides encoding them.
 ACCESSION B063986
 VERSION B063986.1 G1:22609589
 KEYWORDS JP 2001506848-A/10.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 2546)
 AUTHORS Jacobo, K., McCoy, J.M., Lavallie, E.R., Racie, L.A., Merberg, D., Treacy, M., Spaulding, V. and Agostino, M.J.
 TITLE Secreted proteins and polynucleotides encoding them
 JOURNAL Patent: JP 2001506848-A 10 29-MAY-2001;

GENETICS INSTITUTE INC
 PN JP 2001506848-A/10
 PD 29-MAY-2001
 PF 12-DEC-1997 JP 1998525996
 PR 13-DEC-1996 US 08/766263
 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI DAVID MERBERG.
 PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
 C12N15/12, C12N5/10, C07K14/47, A61K38/17
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key Location/Qualifiers.

FEATURES

source

1..2546
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

ORIGIN

Query Match 39.2%; Score 972; DB 6; Length 2546;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1510 ABAAGAAACCAAGGAAATTTGTATGATAGGACAGGTAAAGTGAGGAACATGCAACTGAA 1569
 Db 1 ABAAGAAACCAAGGAAATTTGTATGATAGGACAGGTAAAGTGAGGAACATGCAACTGAA 60
 QY 1570 CAGGAAAAAATCGAAGAGGGATTAGGCCCTAATATAAAAAGCATTTGCCATGTTGATG 1629
 Db 61 CAGGAAAAAATCGAAGAGGGATTAGGCCCTAATATAAAAAGCATTTGCCATGTTGATG 120
 QY 1630 CTGATGCTATTGATGATGTTTGTCTGCCATGTTACCTGGGTCAACAGCATGCTACTCT 1689
 Db 121 CTGATGCTATTGATGATGTTTGTCTGCCATGTTACCTGGGTCAACAGCATGCTACTCT 180
 QY 1690 AGTCCAAAGTGATGCTGGCTCATACATCATGATGGCACCAGGAATATCTTAGATGAT 1749
 Db 181 AGTCCAAAGTGATGCTGGCTCATACATCATGATGGCACCAGGAATATCTTAGATGAT 240
 QY 1750 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGTCT 1809
 Db 241 TTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACGAGTAATGTCT 300
 QY 1810 TGGTGGGATTTAGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGAATAT 1869
 Db 301 TGGTGGGATTTAGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGAATAT 360
 QY 1870 AACACCTGGAATAACAGACCATAGCATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 1929
 Db 361 AACACCTGGAATAACAGACCATAGCATGCTGGTGGGAAAAGCTATGCTTCTAATGAAACA 420
 QY 1930 GCAGCTTAATAATCATGAGACTCTAGATGATAGATTAATGTTGGTATATTTTGGAGGG 1989
 Db 421 GCAGCTTAATAATCATGAGACTCTAGATGATAGATTAATGTTGGTATATTTTGGAGGG 480
 QY 1990 GTTATTGGCTATCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 2049
 Db 481 GTTATTGGCTATCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGAACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 2109
 Db 541 GGAGAACATCCCAAGACATTCGGGAAGTACTATTTTACCCACAGGAGAAATTCCTGT 600
 QY 2110 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATTTGATTAATAAATGTCTACTATAC 2169
 Db 601 GTAGACAAAGCAGGATCCCTACTTTTGGTGAATTTGATTAATAAATGTCTACTATAC 660
 QY 2170 AGATTGGAGAAATGAGCTGGATTTTGTACACCCAGGTTTGGACGACAGCTAAT 2229
 Db 661 AGATTGGAGAAATGAGCTGGATTTTGTACACCCAGGTTTGGACGACAGCTAAT 720
 QY 2230 GCTGAGATTGGAAATTAAGACATTAATAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 2289

/translation="WAEPSPAPESKHSKSLNSPWSGLMALONSRHHGHPGAQCAHKA
AGAAPKPPAPAGLGGSSQAPAGWOSLLSFTILFLAMLAGFSSRLFAVIRPESLHIF
DPWNYRSTHILASHGFYBFLNWFDERAWYPLGRVGVGVVYGLMITAGLTHMLNLT
NITVHRDVCVFLAFSLTISITFLRLBNQAGLLAACFIAIYVGVISRSVAG
SPNEGIALFALQFTYLVKSVKSGVFWMCCLSYFVNVSWAGGVYFLINILPH
FVLLLMORVSEVITAYSTFVIGLILSMQIPRVGPOPIRTSEHMAAGVAFALQAY
AFLOYLRDLTBOEFTLFLGVSLAAGAVELSVLYLTGTGYIAPMSGRVSLWDYGY
AKIHIFIIASVSEHQTTWVSFFDLHLVCTFPFAGLWFCIKNINDERVGA"

ORIGIN

Query Match 51.4%; Score 1276; DB 9; Length 1664;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGGAGCCCTCGCGCCCGGAGAGCAAGCACTGCTCCCTCAACTCGTCCCGTGG 60
DB |||||
QY 28 ATGCGGAGCCCTCGCGCCCGGAGAGCAAGCAAGTGTCTCCCTCAACTCGTCCCGTGG 87
DB |||||
QY 61 AGTGGCCTCATGCGCCCTGGGAAACAGCCGCGCACCGCCACCGCGCCCGGCGCCAGTGC 120
DB |||||
QY 88 AGTGGCCTCATGCGCCCTGGGAAACAGCCGCGCACCGCCACCGCGCCCGGCGCCAGTGC 147
DB |||||
QY 121 GCGCACAGCG 180
DB |||||
QY 148 GCGCACAGCG 207
DB |||||
QY 181 GGGCTGTGCGAGCG 240
DB |||||
QY 208 GGGCTGTGCGAGCG 267
DB |||||
QY 241 TGGCTTCCCGGCTTCAGCTCGCGCGCTTCGCGGCTCCGCGCTTCGCGGAGCATCATCCAC 300
DB |||||
QY 268 TGGCTTCCCGGCTTCAGCTCGCGCGCTTCGCGGCTTCGCGGCTTCGCGGAGCATCATCCAC 327
DB |||||
QY 301 GAGTTCCACCGCGGTTTAACTATAGATCAACACATCATCTTGTGCTCATCTGCGGTTCTAT 360
DB |||||
QY 328 GAGTTCCACCGCGGTTTAACTATAGATCAACACATCATCTTGTGCTCATCTGCGGTTCTAT 387
DB |||||
QY 361 GAATTTTAAATGTTTGAATGAAGAGCATGGTATCCACTAGGAGATAGTAGGTTGGT 420
DB |||||
QY 388 GAATTTTAAATGTTTGAATGAAGAGCATGGTATCCACTAGGAGATAGTAGGTTGGT 447
DB |||||
QY 421 ACTGTTTACCAGGTTGATGATPACCGCTGCGCTTATTCATTTGATTTTAAATACATTG 480
DB |||||
QY 448 ACTGTTTACCAGGTTGATGATPACCGCTGCGCTTATTCATTTGATTTTAAATACATTG 507
DB |||||
QY 481 AACATCTCTTACATAGAGAGCATGATGTTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 540
DB |||||
QY 508 AACATACTGTTACATAGAGAGCATGATGTTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 567
DB |||||
QY 541 ACATCTATCTACTTCTCTGCTTACAGAGAACTTTGGAACCAAGGAGCAGGACTTTTA 600
DB |||||
QY 568 ACATCTATCTACTTCTCTGCTTACAGAGAACTTTGGAACCAAGGAGCAGGACTTTTA 627
DB |||||
QY 601 GCTGCTTGTGTTTATGCTATGTGACAGAGCATATATCTGCTGCTGCTGCTGCTGCTGCTT 660
DB |||||
QY 628 GCTGCTTGTGTTTATGCTATGTGACAGAGCATATATCTGCTGCTGCTGCTGCTGCTGCTT 687
DB |||||
QY 661 GATAATGAAGGCTATGCTATTTTGTCACTTCAGTTCACATATATTTATGCGTAAATCT 720
DB |||||
QY 688 GATAATGAAGGCTATGCTATTTTGTCACTTCAGTTCACATATATTTATGCGTAAATCT 747
DB |||||
QY 721 GTAAATCTGGGTCAGTTTGTGCAAGTGTGCTGCTGCTTATCTATTTCTATATGTC 780
DB |||||
QY 748 GTAAATCTGGGTCAGTTTGTGCAAGTGTGCTGCTGCTTATCTATTTCTATATGTC 807
DB |||||
QY 781 TCTGCTTGGGTCAGTTTGTGCAAGTGTGCTGCTGCTTATCTATTTCTATATGTC 840
DB |||||
QY 808 TCTGCTTGGGTCAGTTTGTGCAAGTGTGCTGCTGCTTATCTATTTCTATATGTC 867
DB |||||
QY 841 TTACTGATGAGAGATACACAAAGAGTTCATAGCATATAGCATTTCTTACATCTG 900
DB |||||
QY 868 TTACTGATGAGAGATACACAAAGAGTTCATAGCATATAGCATTTCTTACATCTG 927
DB |||||

QY 901 GGTTAATATTATCAATGCAGATACCTTTTGTGGGATTCAGCCATCAGACAACTGAA 960
DB |||||
QY 928 GGTTAATATTATCAATGCAGATACCTTTTGTGGGATTCAGCCATCAGACAACTGAA 987
DB |||||
QY 961 CACATGGCAGCTGCAAGGTGCTTTGCAATTCGCAAGCTTATGCTTTCTTGCAGTATCTG 1020
DB |||||
QY 988 CACATGGCAGCTGCAAGGTGCTTTGCAATTCGCAAGCTTATGCTTTCTTGCAGTATCTG 1047
DB |||||
QY 1021 AGAGACCATTAACAAACAAAGAGTTCAGACCCCTTTTCTTTTGGGTGATCAGTAGCT 1080
DB |||||
QY 1048 AGAGACCATTAACAAACAAAGAGTTCAGACCCCTTTTCTTTTGGGTGATCAGTAGCT 1107
DB |||||
QY 1081 GCAGGTGCTGCTGCTTCTTAGTGTCTATCTATTGACTTATACAGGTTACATGACCATGG 1140
DB |||||
QY 1108 GCAGGTGCTGCTGCTTCTTAGTGTCTATCTATTGACTTATACAGGTTACATGACCATGG 1167
DB |||||
QY 1141 AGTGCAGGTTTATCTATTGCGGATCTGCGGTATCGAAATATACATCCCAATTAT 1200
DB |||||
QY 1168 AGTGCAGGTTTATCTATTGCGGATCTGCGGTATCGAAATATACATCCCAATTAT 1227
DB |||||
QY 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGTGATCTACATATT 1260
DB |||||
QY 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTGTGATCTACATATT 1287
DB |||||
QY 1261 CTTGTATGATCTTCCAGAGCGCTTGGTTCGCAATCAAAATATCAACGATGAAGA 1320
DB |||||
QY 1288 CTTGTATGATCTTCCAGAGCGCTTGGTTCGCAATCAAAATATCAACGATGAAGA 1347
DB |||||
QY 1321 GTATTG 1327
DB |||||
QY 1348 GTATTG 1354
DB |||||

RESULT 6
AX099510
LOCUS 2546 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 150 from Patent WO0119988.
ACCESSION AX099510
VERSION AX099510.1 GI:13538588
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jacobs K., McCoy J.M., Lavallie E.R., Collins-Racie L.A., Evans C.,
Merberg D., Treacy M., Bowman M.R., Spaulding V. and Agostino M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 150 22-MAR-2001;
Genetics Institute, Inc. (US)
LOCATION/Qualifiers
FEATURES
source
1..2546
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 39.2%; Score 972; DB 6; Length 2546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 AAAAGAAACCAAGGAAATTTGTATGAAGGAGGTAAGGAGGAAACATGCAACTGAA 1569
DB |||||
QY 1570 CAGGAAAAAATGAAAGAGGATAGGCCCTTAATAAAAAAGCATTTGCACCATTTGATG 1629
DB |||||
QY 1630 CAGGAAAAAATGAAAGAGGATAGGCCCTTAATAAAAAAGCATTTGCACCATTTGATG 1689
DB |||||
QY 1630 CAGGAAAAAATGAAAGAGGATAGGCCCTTAATAAAAAAGCATTTGCACCATTTGATG 1700
DB |||||

Query Match 51.4%; Score 1276; DB 6; Length 1664;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGG 60
DB 28 ATGCGGAGCCCTCGGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGG 87
QY 61 AGTGGCTCATGGCCCTGGGAAACAGCGGSCACGCGGCAACGCGGGCCGAGTGC 120
DB 88 AGTGGCTCATGGCCCTGGGAAACAGCGGSCACGCGGCAACGCGGGCCGAGTGC 147
QY 121 GCGCAAGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 148 GCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 207
QY 191 GGGCTGTCGACCGCGCTGGGTGGAGTGCCTCTCTCTCTCTCTCTCTCTCT 240
DB 208 GGGCTGTCGACCGCGCTGGGTGGAGTGCCTCTCTCTCTCTCTCTCTCTCT 267
QY 241 TGGCTTGGCGGCTTCAGCTCGCGCTCTTGGCGGTCATCGCTTGGAAAGCATTCAC 300
DB 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTGGCGGTCATCGCTTGGAAAGCATTCAC 327
QY 301 GAGTTGCGACCGCTGGTTAACTAGATCAACATCATCTTGCATCTCATGCTTCTAT 360
DB 328 GAGTTGCGACCGCTGGTTAACTAGATCAACATCATCTTGCATCTCATGCTTCTAT 387
QY 361 GAAATTTTAAATGGTTTGAAGAGAGCATGGTATCCACTAGGAGAAATAGTAGTGGT 420
DB 388 GAAATTTTAAATGGTTTGAAGAGAGCATGGTATCCACTAGGAGAAATAGTAGTGGT 447
QY 421 ACTGTTTACCGAGGTTGATGATTAACCGCTGGCTTATTCATGGATTTAAATACATG 480
DB 448 ACTGTTTACCGAGGTTGATGATTAACCGCTGGCTTATTCATGGATTTAAATACATG 507
QY 481 AACATACTGTTCATAGAAGAGCATGTGTGCTTCTTGCACCAACTTTTACGGCGCTT 540
DB 508 AACATACTGTTCATAGAAGAGCATGTGTGCTTCTTGCACCAACTTTTACGGCGCTT 567
QY 541 ACATCTATATCTCTTCTGCTTACAGAGAGACTTTGGAAACCAAGGACGAGCTTTTA 600
DB 568 ACATCTATATCTCTTCTGCTTACAGAGAGACTTTGGAAACCAAGGACGAGCTTTTA 627
QY 601 GCTGCTGTTTTATGCTATTTGACAGGCTACATATCTGCTGCTAGTGCATGCTTT 660
DB 628 GCTGCTGTTTTATGCTATTTGACAGGCTACATATCTGCTGCTAGTGCATGCTTT 687
QY 661 GATATGAAGCATTGCTATTTTGCACCTTCAGTTCACATCTATTTATGGCTAAATCT 720
DB 688 GATATGAAGCATTGCTATTTTGCACCTTCAGTTCACATCTATTTATGGCTAAATCT 747
QY 721 GTAAAACTGGGTGAGTTTGGACAAATGTGCTGCTTATCTTATCTTATATGCTC 780
DB 748 GTAAAACTGGGTGAGTTTGGACAAATGTGCTGCTTATCTTATCTTATATGCTC 807
QY 781 TCTGCTGGGTGGTGTATGATTTATCATCAATCTTATTCCTGCTGCTGCTGCTGCTG 840
DB 808 TCTGCTGGGTGGTGTATGATTTATCATCAATCTTATTCCTGCTGCTGCTGCTGCTG 867
QY 841 TTACTGATGCAAGATACAGCAAGAGTCTACATAGCATATAGCATCTTCTACATGTTG 900
DB 868 TTACTGATGCAAGATACAGCAAGAGTCTACATAGCATATAGCATCTTCTACATGTTG 927
QY 901 GGTTTTAAATTTATCAATGCAATACCTTTTGTGGGATTCAGCAATCAGAACAGTGA 960
DB 928 GGTTTTAAATTTATCAATGCAATACCTTTTGTGGGATTCAGCAATCAGAACAGTGA 987
QY 961 CACATGGAGCTGAGGTGCTTGTGATTCCTGCAAGCTATATGCTTCTTGCAGTATCTG 1020
DB 988 CACATGGAGCTGAGGTGCTTGTGATTCCTGCAAGCTATATGCTTCTTGCAGTATCTG 1047
QY 1021 AGAGACCGGATTAACAAAAAAGAGTTCAGACCCCTTTCTTTTGGGTGCTCACTAGCT 1080

1048 AGAGACCGGATTAACAAAAAAGAGTTCAGACCCCTTTCTTTTGGGTGCTCACTAGCT 1107
1081 GCAGGTGCTGCTGCTTCTTCTAGTGTCACTATTGACTTATACAGGTACATGACCATGG 1140
1108 GCAGGTGCTGCTGCTTCTTCTAGTGTCACTATTGACTTATACAGGTACATGACCATGG 1167
1141 AGTGCAGGTTTTATTGATTTGGGATCTGGGATCGGATCGGATCGGATCGGATCGGATCGGAT 1200
1168 AGTGCAGGTTTTATTGATTTGGGATCTGGGATCGGATCGGATCGGATCGGATCGGATCGGAT 1227
1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTGATCTACATATT 1260
1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTGATCTACATATT 1287
1261 CTTGTATGCTCTCCAGCAGCGCTTGTGCTGCTCAAAAATATCAACGATGAAGA 1320
1288 CTTGTATGCTCTCCAGCAGCGCTTGTGCTGCTCAAAAATATCAACGATGAAGA 1347
1321 GTATTG 1327
1348 GTATTG 1354

RESULT 5
AK074587
LOCUS
DEFINITION
Homo sapiens cDNA FLJ90106 f1s, clone HEMBA1006430, weakly similar
to Human putative transmembrane protein precursor (B5) mRNA.
ACCESSION
AK074587
VERSION
AK074587.1 GI:22760122
KEYWORDS
oligo capping; f1s (full insert sequence).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
Kojima,S., Nagahara,K., Masuho,Y., Ono,T., Okano,K., Yoshikawa,Y.,
Aotaka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
Ninomiya,K.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1664)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).
FEATURES
Location/Qualifiers
1..1664
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEMBA1006430"
/tissue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/notes="cloning vector: pME18SFL3"
28..1382
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11076.1"
/db_xref="GI:22760123"

CDS

DPNFNVRSTHLLASHGFYEFLLNPFDERAWYPLGRIVGGTVVPLMITAGLHMLNTL
 NITVHRDVCFLAFTFSLTISITFLTRLMQAGGLAAFCIAIUPGYISRSVAG
 SFDEGIAIFALQFTYFLWVKSVGTWVWCCCLSYFVMSAWGVFIINLIPLH
 VFVLLLMORVSRVIAVSTFVGLIISMOIPFVGFOPRTSHMAAGVFAILOAY
 AFLOYLDRLTQEBQTFPLGVSLAAGAVLSVLYLTGYIAPWGRSVSLWDIGY
 AKIHIFLIASVSEHQPTWVSFFDLHLVCTFPAGLWFCIKINDERVGA"

ORIGIN

Query Match 51.4%; Score 1276; DB 6; Length 1664;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGGAGCCCTCGGCCCGGAGACGACAGTGTCTCACTCGTCCCGGTGG 60
 Db |||||
 Qy 28 ATGCGGAGCCCTCGGCCCGGAGACGACAGTGTCTCACTCGTCCCGGTGG 87
 Db |||||
 Qy 61 AGTGGCTCATGCTCGTGGAAACAGCGCGACCGCCACCGCGCGCGCGCGTGC 120
 Db |||||
 Qy 88 AGTGGCTCATGCTCGTGGAAACAGCGCGACCGCCACCGCGCGCGCGCGTGC 147
 Db |||||
 Qy 121 GGCACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 180
 Db |||||
 Qy 148 GCGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGC 207
 Db |||||
 Qy 181 GGGCTGTCGACGCGCGTGGTGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 240
 Db |||||
 Qy 208 GGGCTGTCGACGCGCGTGGTGGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 267
 Db |||||
 Qy 241 TGGCTGTCGCGCTGAGTGGCGCTCTTGGCGCTCTTGGCGCTCTTGGCGCTCT 300
 Db |||||
 Qy 268 TGGCTGTCGCGCTGAGTGGCGCTCTTGGCGCTCTTGGCGCTCTTGGCGCTCT 327
 Db |||||
 Qy 301 GAGTTCGACCGCGTGGTAACTATAGATCAACACATCATCTTGTCTCATCTGCT 360
 Db |||||
 Qy 328 GAGTTCGACCGCGTGGTAACTATAGATCAACACATCATCTTGTCTCATCTGCT 387
 Db |||||
 Qy 361 GAATTTTAAATGGTGTGATGAAGACGATGGTATCCAGTGAAGATAGTAGTGGT 420
 Db |||||
 Qy 388 GAATTTTAAATGGTGTGATGAAGACGATGGTATCCAGTGAAGATAGTAGTGGT 447
 Db |||||
 Qy 421 ACTGTTTACCGAGGTGATGATPACCGCTGCGCTTATTCATGATGATTTAAATCAT 480
 Db |||||
 Qy 448 ACTGTTTACCGAGGTGATGATPACCGCTGCGCTTATTCATGATGATTTAAATCAT 507
 Db |||||
 Qy 481 AACATRACTCTCACAATGAGACGATGTGTCTTCTTCTTCTTCTTCTTCTTCTT 540
 Db |||||
 Qy 508 AACATRACTCTCACAATGAGACGATGTGTCTTCTTCTTCTTCTTCTTCTTCTT 567
 Db |||||
 Qy 541 ACATCTATATCTACTTCTCTGTTTACAGAGAACTTTGGAACCAAGGACGAGCTTTTA 600
 Db |||||
 Qy 568 ACATCTATATCTACTTCTCTGTTTACAGAGAACTTTGGAACCAAGGACGAGCTTTTA 627
 Db |||||
 Qy 601 GCTGCTTGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 660
 Db |||||
 Qy 628 GCTGCTTGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 687
 Db |||||
 Qy 661 GATAATGAAGGCAATGCTATTTTGGCACTTCAGTTTACATATCTATTTATGGTAAATCT 720
 Db |||||
 Qy 688 GATAATGAAGGCAATGCTATTTTGGCACTTCAGTTTACATATCTATTTATGGTAAATCT 747
 Db |||||
 Qy 721 GTAAARACTGGTCAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGA 780
 Db |||||
 Qy 748 GTAAARACTGGTCAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGA 807
 Db |||||
 Qy 781 TCTGCTTGGGTGGTATGATTTATCATCAATCTTATTCATCTCATGATTTGTTGTTG 840
 Db |||||
 Qy 808 TCTGCTTGGGTGGTATGATTTATCATCAATCTTATTCATCTCATGATTTGTTGTTG 867
 Db |||||
 Qy 841 TTACTGATGAGAGATACACAAAGAGTCTACATAGCATATAGCATTTTCTACATTTG 900
 Db |||||
 Qy 868 TTACTGATGAGAGATACACAAAGAGTCTACATAGCATATAGCATTTTCTACATTTG 927
 Db |||||
 Qy 901 GGTTTAATATATCAATGACAGATACCTTTTGTGGGATTCACGCAATCAGAACAGTGAA 960
 Db |||||

Db 928 GGTTTAATATATCAATGACAGATACCTTTTGTGGATTCAGCAATCAGAACAGTGAA 987
 Qy 961 CACATGGCAGCTGCGAGGTGCTTTTGCATTTGCTGCAAGCTTATCTTTCTTGCAGTATCTG 1020
 Db 988 CACATGGCAGCTGCGAGGTGCTTTTGCATTTGCTGCAAGCTTATCTTTCTTGCAGTATCTG 1047
 Qy 1021 AGAGACCGATTAACAAACAAAGATTCAGACCCCTTTTCTTTTGGTGTATCAGTACCT 1080
 Db 1048 AGAGACCGATTAACAAACAAAGATTCAGACCCCTTTTCTTTTGGTGTATCAGTACCT 1107
 Qy 1081 GCAGTGTCTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
 Db 1108 GCAGTGTCTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1167
 Qy 1141 AGTGCAGCTTTTATTCATTTGCGGATCTGGGTATGCAAAATACACATTCCTCAATATT 1200
 Db 1168 AGTGCAGCTTTTATTCATTTGCGGATCTGGGTATGCAAAATACACATTCCTCAATATT 1227
 Qy 1201 GCATCAGTGTCTGAGCATCAACCTACAGCTTGGGTGCTTTCTTCTTCTTCTTCTTCTTCTTCT 1260
 Db 1228 GCATCAGTGTCTGAGCATCAACCTACAGCTTGGGTGCTTTCTTCTTCTTCTTCTTCTTCTTCT 1287
 Qy 1261 CTTGTATGATCTTCCAGCAGGCTTTGGTCTGCAATCAAAATATCAACGATGAAGA 1320
 Db 1288 CTTGTATGATCTTCCAGCAGGCTTTGGTCTGCAATCAAAATATCAACGATGAAGA 1347
 Qy 1321 GTATTGT 1327
 Db 1348 GTATTGT 1354

RESULT 4

BD127193
 LOCUS 1664 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127193
 VERSION BD127193.1 GI:23222138
 KEYWORDS JP 2002017375-A/2624.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2624 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2624
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOGA
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10, C12P21/02, C12Q1/68/ C12P21/08 G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof PH Key

FT CDS Location/Qualifiers
 1..1664
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

FEATURES
 source
 1..1664

ORIGIN

Db 781 TCTGCTGGGTTGTTATTTATCATCAATCTTATCCACTGCTATGTTGTTG 840
QY 841 TTACTGATGACAGATACAGAAAGAGTCTCATAGCATATAGCACTTTTACATTTG 900
Db 841 TTACTGATGACAGATACAGAAAGAGTCTCATAGCATATAGCACTTTTACATTTG 900
QY 901 GGTTTAATATTATCAATGACAGATACCTTTTGTGGGATTCAGCAATACAGAAAGTGA 960
Db 901 GGTTTAATATTATCAATGACAGATACCTTTTGTGGGATTCAGCAATACAGAAAGTGA 960
QY 961 CACATGGAGCTGACAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020
Db 961 CACATGGAGCTGACAGTGTCTTTGCAATGCTGCAAGCTTATGCTTTCTTGCAATCTG 1020
QY 1021 AGAGACCGATTAAACAAACAGAGTCTCCAGACCTTTTCTTTTGGGTATCACTAGCT 1080
Db 1021 AGAGACCGATTAAACAAACAGAGTCTCCAGACCTTTTCTTTTGGGTATCACTAGCT 1080
QY 1081 GCAGGTGCTGTCTTCTTAGTGCTATCTATTGACTTATACAGTTTACATTCACCATGG 1140
Db 1081 GCAGGTGCTGTCTTCTTAGTGCTATCTATTGACTTATACAGTTTACATTCACCATGG 1140
QY 1141 AGTGGCAGGTTTATTCATTGTGGGATCTGCTGCTATGCAAAATACACATTCCTAATTT 1200
Db 1141 AGTGGCAGGTTTATTCATTGTGGGATCTGCTGCTATGCAAAATACACATTCCTAATTT 1200
QY 1201 GCATCAGTCTGACGATCAACCTGACCTGCTGCTGCTCTTCTTTGATCTACATTT 1260
Db 1201 GCATCAGTCTGACGATCAACCTGACCTGCTGCTGCTCTTCTTTGATCTACATTT 1260
QY 1261 CTTGTATGATCTTCCAGCAGGCTTTTGGTCTGATCAAAATATCAACGATGAAAGA 1320
Db 1261 CTTGTATGATCTTCCAGCAGGCTTTTGGTCTGATCAAAATATCAACGATGAAAGA 1320
QY 1321 GTATTTGCTGCTATATGCAATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
Db 1321 GTATTTGCTGCTATATGCAATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
QY 1381 ATGTTGATCTTGAATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 1381 ATGTTGATCTTGAATCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1441 GAGCATTATTTGGGGATGATCATGAAGGGAATCCACCTGCTGGAGCAGCTGCT 1500
Db 1441 GAGCATTATTTGGGGATGATCATGAAGGGAATCCACCTGCTGGAGCAGCTGCT 1500
QY 1501 GAGGATGACAAAGAAACCAAGGAAATTTCTATGATAGGAGGTTAAAGTGGAGAAACAT 1560
Db 1501 GAGGATGACAAAGAAACCAAGGAAATTTCTATGATAGGAGGTTAAAGTGGAGAAACAT 1560
QY 1561 GCACTGAAACAGGAAACCAAGGAAATTTAGGCTTAAATATAAAGGATTTGTCACC 1620
Db 1561 GCACTGAAACAGGAAACCAAGGAAATTTAGGCTTAAATATAAAGGATTTGTCACC 1620
QY 1621 ATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1621 ATGTTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 GCCTACTCTAGTCCAAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Db 1681 GCCTACTCTAGTCCAAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
QY 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAAATGACAGA 1800
Db 1741 TTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAAATGACAGA 1800
QY 1801 GTATGCTCTGCTGGGATTTAGGCTATCAGATAGCTGGAATGCTAATGAACTACGTTG 1860
Db 1801 GTATGCTCTGCTGGGATTTAGGCTATCAGATAGCTGGAATGCTAATGAACTACGTTG 1860
QY 1861 GTGGATATACACTGGAATTAACAGCCATAGCTGCTGGGAAAGCTATGCTTCT 1920

Db 1861 GTGGATATTAACACCTGGAATAACAGCCATAGCACTGCTGGGAAAGCTATGCTTCT 1920
QY 1921 AATGAAACAGCAGCTATAAAATCATGAGGACTCTAGATGATGATGATGATGATGAT 1980
Db 1921 AATGAAACAGCAGCTATAAAATCATGAGGACTCTAGATGATGATGATGATGATGAT 1980
QY 1981 TTTGGAGGGTTATTTGGCTATTTCTGCTGATCATCAAAATTTCTCTGATGTTAGG 2040
Db 1981 TTTGGAGGGTTATTTGGCTATTTCTGCTGATCATCAAAATTTCTCTGATGTTAGG 2040
QY 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2100
Db 2041 ATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAAGTGAATTTTACCCCAAGGGA 2100
QY 2101 GAATTCCTGTTAGACAAAGCAGGATCCCTACTTTTGTGATTCCTTATGATGATGAT 2160
Db 2101 GAATTCCTGTTAGACAAAGCAGGATCCCTACTTTTGTGATTCCTTATGATGATGAT 2160
QY 2161 TCATCTACAGATTTGGGAGAAATGCACTGATTTTCTGACACCCAGCTTTTGACCGA 2220
Db 2161 TCATCTACAGATTTGGGAGAAATGCACTGATTTTCTGACACCCAGCTTTTGACCGA 2220
QY 2221 ACACGTAATGCTGAGATTTGGAATTAAGGACATTTAAATTTCAAAATTTTGAAGAGCTTT 2280
Db 2221 ACACGTAATGCTGAGATTTGGAATTAAGGACATTTAAATTTCAAAATTTTGAAGAGCTTT 2280
QY 2281 ACATCAGACACTGCTGTTAGGATATATTAAGTAAAGCCTGATACACCCAGCTTTGACCGA 2340
Db 2281 ACATCAGACACTGCTGTTAGGATATATTAAGTAAAGCCTGATACACCCAGCTTTGACCGA 2340
QY 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTTGTCAAGAAG 2400
Db 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTTGTCAAGAAG 2400
QY 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAA 2460
Db 2401 ACTACCAAAAGGAGCGTGGCTACATTAATAAATAGCTGTTTAAAGAAAGCAAGAA 2460
QY 2461 ATATCTAAGAGACTGTTTAA 2481
Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 3

CQ782554

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

CQ782554 1664 bp DNA linear PAT 17-MAR-2004
Sequence 2694 from Patent EP1396543.

CQ782554.1 GI:45502338

KEYWORDS

ORGANISM Homo sapiens (human)

REFERENCE 1

Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and Koga, H.

Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,

Primer for synthesizing full length cDNA clones and their use

Patent: EP 1396543-A 2694 10-MAR-2004;

Research Association for Biotechnology (JP)

Location/Qualifiers

1. 1664

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

28. 1362

/note="unnamed protein product"

/codon_start=1

/protein_id="CAF95831.1"

/db_xref="GI:45502338"

/translation="MAEFSAPESKHKSLSNPSWGLMGLNRSRHGHGHPGACAKHA

AGGAAPPKPAAGLSGGLSPAGWQSLLSFTILFLAWLAFGLSRLFAVIRFESIIHEF

QY 2101 GAATTCGGTGTAGCAAGCAGGATCCCTACTTCTTGAATTCCTTATGTAATAATG 2160
 DB 2101 GAATTCGGTGTAGCAAGCAGGATCCCTACTTCTTGAATTCCTTATGTAATAATG 2160
 QY 2161 TCATACACAGATTGGAGAAATCGAGCTGGATTTTCGTACACCCCGAGTTTGAACGA 2220
 DB 2161 TCATACACAGATTGGAGAAATCGAGCTGGATTTTCGTACACCCCGAGTTTGAACGA 2220
 QY 2221 ACAGTAAATCTGAGATTGGAATAAGGACATTAATTCACACATTTGGAAGAGCCCTTT 2280
 DB 2221 ACAGTAAATCTGAGATTGGAATAAGGACATTAATTCACACATTTGGAAGAGCCCTTT 2280
 QY 2281 ACATCAGAACACTGGCTGTGTAGATATATAAAGTAAAGACCTGTATAACAGGAGACA 2340
 DB 2281 ACATCAGAACACTGGCTGTGTAGATATATAAAGTAAAGACCTGTATAACAGGAGACA 2340
 QY 2341 TTAGATCACAACCTCGAGTCACCAATTTTCCCAAAACAGAGTATTTGTCAAAGAG 2400
 DB 2341 TTAGATCACAACCTCGAGTCACCAATTTTCCCAAAACAGAGTATTTGTCAAAGAG 2400
 QY 2401 ACTACCAAAAGAGAGCGTGTACATTAATAATAGCTGGTTTAAAGAAAGGCAAGAAA 2460
 DB 2401 ACTACCAAAAGAGAGCGTGTACATTAATAATAGCTGGTTTAAAGAAAGGCAAGAAA 2460
 QY 2461 ATATCTAAGAGAGACTGTTTAA 2481
 DB 2461 ATATCTAAGAGAGACTGTTTAA 2481

RESULT 2
 AY074880 2481 bp mRNA linear PRI 10-DEC-2002
 LOCUS Homo sapiens source of immunodominant MHC-associated peptides
 DEFINITION (SIMP) mRNA, complete cds.
 ACCESSION AY074880
 VERSION AY074880.1 GI:19879588
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 2481)
 AUTHORS McBride, K., Baron, C., Picard, S., Martin, S., Boismenu, D., Bell, A.,
 Bergeron, J., and Perreault, C.
 TITLE The model B6dom1 minor histocompatibility antigen is encoded by a
 mouse homolog of the yeast Sit3 gene
 JOURNAL Immunogenetics 54 (8), 562-569 (2002)
 MEDLINE 2326278
 PUBMED 12439619
 REFERENCE 2 (bases 1 to 2481)
 AUTHORS McBride, K. and Perreault, C.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-2002) Molecular Biology Group, Compatible, 6100
 Royalmount, Montreal, QC H4P 2R2, Canada

FEATURES
 source
 1. 2481
 /organism="Homo sapiens"
 /note="similar to Saccharomyces cerevisiae OSTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /db_xref="taxon:9606"
 /chromosome="3"
 1. 2481
 /gene="SIMP"
 1. 2481
 /gene="SIMP"
 1. 2481
 /note="similar to Saccharomyces cerevisiae OSTT3 and Mus
 musculus protein encoded by GenBank Accession Number
 AK018758; source of minor histocompatibility antigen(s)"
 /codon_start=1
 /product="source of immunodominant MHC-associated
 peptides"
 /protein_id="AA171884.1"
 /db_xref="GI:19879589"
 /translation="MAEPAPESKHKSSLSNSPGLMALGNSRHGHGPGCAQAHKA

AGGAAPKAPAGLGGGLSOPACWQSLISFTILFLAWLAFSSRLPAVIRESIHEP
 DPWNYSTHLLASHGFEYELNWFDRANYPBLGRIVGCTVYELMAGLJHMLNLT
 NITVHSDVCVFLAFTPSGLTISITELLRELWNGGLAAACFAITVPGIISRSVAG
 SFDNEGIAIFALQFTYILWVKSVKTSVFTWCCCLSYFYFVMSAWGIVFIINLPLH
 VFVLLLRYSKRKYIAYSTFYVGLILSMQIPFVGFQPIRTSEHMAAGFALQAY
 AFLGLDRLTQBFOTFFLGLSLAAGFLSVLYLTVTYIAWMSRFYSLWMDTG
 AKIHIPITAGVSHQPTTWVSPFFDLHLVCTFPAGLWFCIKINDEKRVFALLAISA
 VYFAGVVRMLTLTPVCMLSAIFSNVFEHLVGDMDKRENNPVEDSDSDKRNQO
 NLXKACKVRKHATEOEKTEEGGPNIKSIVTMLMLMLMFAVCTVWVNTNAYSSPS
 VLASVNHDCRTNLILODPBEAYFELPONTDEHARVMSWDYGYQJAGVANTLVDNN
 TANNSHALVKGAMSSNETAAYKIMETLVDYVLYIFGSGVIGYSGDDINKELWVRIA
 EGEHPKIDRESDIYFTPGSEFRVDKAGSPILLNCLMTKMSIYRFGSEQLDRIFFPGFDR
 TRNAIEGNKIDIKFHLKEEAFPSHMLVRIYKVKAPDNRETLDHKPRVTINIPPKQKLS
 KKTTTKRGYIKNKLVEKKGKISKKTV"
 ORIGIN
 Query Match 100.0%; Score 2481; DB 9; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCGGAGCCCTCGGCGCCCGGAGAGCAACAAGTGTCTCTCACTCGTCCCGTGG 60
 DB 1 ATGCGGAGCCCTCGGCGCCCGGAGAGCAACAAGTGTCTCTCACTCGTCCCGTGG 60
 QY 61 AGTGGCCTCATGCGCCTGGGAAACAGCCGCGACCGCCACCGGGCCCGGCGCCAGTGC 120
 DB 61 AGTGGCCTCATGCGCCTGGGAAACAGCCGCGACCGCCACCGGGCCCGGCGCCAGTGC 120
 QY 121 GCGCAAAAGCG 180
 DB 121 GCGCAAAAGCG 180
 QY 181 GGGCTGTGCGACCGCGGTGGGTGGCAGTCTCTCTCTCTCAACATCTCTCTCTGCCC 240
 DB 181 GGGCTGTGCGACCGCGGTGGGTGGCAGTCTCTCTCTCTCTCAACATCTCTCTCTGCCC 240
 QY 241 TGGCTTGGCGGCTTCACTGCGGCTTTCGCGCTCATGCGGTCGGAAGCATCATCCAC 300
 DB 241 TGGCTTGGCGGCTTCACTGCGGCTTTCGCGCTCATGCGGTCGGAAGCATCATCCAC 300
 QY 301 GAGTTCGACCGCGTGGTTAACTATAGATCAACATCATCTTGCATCTCATGGTCTTAT 360
 DB 301 GAGTTCGACCGCGTGGTTAACTATAGATCAACATCATCTTGCATCTCATGGTCTTAT 360
 QY 361 GAATTTTAAATGGTTTGAAGAGCATGTATCCACTAGGAGATAGTAGTGGT 420
 DB 361 GAATTTTAAATGGTTTGAAGAGCATGTATCCACTAGGAGATAGTAGTGGT 420
 QY 421 ACTGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCTTGGATTATAAATCATTTG 480
 DB 421 ACTGTTTACCCAGGTTGATGATAACCGCTGGCTTATTCTTGGATTATAAATCATTTG 480
 QY 481 ACATTAATCTGTTACATAAGAGCATGTATGCTTCTTGCACCACTTTTACCGGCTT 540
 DB 481 ACATTAATCTGTTACATAAGAGCATGTATGCTTCTTGCACCACTTTTACCGGCTT 540
 QY 541 ACATCTATATCTACTTTCTCTTACAGAGAACTTTTGGAAACCAAGGAGGAGACTTTTA 600
 DB 541 ACATCTATATCTACTTTCTCTTACAGAGAACTTTTGGAAACCAAGGAGGAGACTTTTA 600
 QY 601 GCTGCTGTTTATTGCTTATTGTACAGGCTCATATCTCGGTAGTAGTGGTCCCTTT 660
 DB 601 GCTGCTGTTTATTGCTTATTGTACAGGCTCATATCTCGGTAGTAGTGGTCCCTTT 660
 QY 661 GATAATGAAGCATGCTATTTTGTCACTTCAGTTCACATCTATTTATGGTAAATCT 720
 DB 661 GATAATGAAGCATGCTATTTTGTCACTTCAGTTCACATCTATTTATGGTAAATCT 720
 QY 721 GTAAATACTGGGTGAGTTTGTGACAAATGTGCTGCTTATCTTATTTCTATATGTC 780
 DB 721 GTAAATACTGGGTGAGTTTGTGACAAATGTGCTGCTTATCTTATTTCTATATGTC 780
 QY 781 TCTGCTGGGTGGTTATGATTAATTAATCAATCTTATCCACTGATGATGTTGTTG 840

ORIGIN

Query Match 100.0%; Score 2481; DB 6; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCTCCCTCAACTCGTCCCGTGG 60
 DB 1 ATGGCGAGCCCTCGGCGCGGAGCAAGCAAGTCTCCCTCAACTCGTCCCGTGG 60

QY 61 AGTGGCTCATGGCTCGGAGCAAGCAAGTCTCCCTCAACTCGTCCCGTGG 120
 DB 61 AGTGGCTCATGGCTCGGAGCAAGCAAGTCTCCCTCAACTCGTCCCGTGG 120

QY 121 GCGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 DB 121 GCGCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180

QY 181 GGGCTGTCGAGCGGCTGGGTGGCACTGCTCTCTCTCTCAACATCTCTCTGGCC 240
 DB 181 GGGCTGTCGAGCGGCTGGGTGGCACTGCTCTCTCTCTCAACATCTCTCTGGCC 240

QY 241 TGGCTTGGCGGCTTCAAGTGGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGG 300
 DB 241 TGGCTTGGCGGCTTCAAGTGGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGG 300

QY 301 GAGTTCGAGCGGCTTCAATAGATCAACACATCTCTCTCTCTCTCTCTCTCTAT 360
 DB 301 GAGTTCGAGCGGCTTCAATAGATCAACACATCTCTCTCTCTCTCTCTCTAT 360

QY 361 GAATTTTAAATGGTTTGAAGAGAGATGATGATGATGATGATGATGATGATG 420
 DB 361 GAATTTTAAATGGTTTGAAGAGAGATGATGATGATGATGATGATGATGATG 420

QY 421 ACTGTTTACCAAGGCTTGAATGATGATGATGATGATGATGATGATGATGATG 480
 DB 421 ACTGTTTACCAAGGCTTGAATGATGATGATGATGATGATGATGATGATGATG 480

QY 481 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 DB 481 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540

QY 541 ACATCTATATCT 600
 DB 541 ACATCTATATCT 600

QY 601 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 601 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660

QY 661 GATATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 DB 661 GATATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

QY 721 GTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
 DB 721 GTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

QY 781 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 DB 781 TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

QY 841 TTACTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 841 TTACTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

QY 901 GGTTTAATATATCAATGAGATACCTTTTGTGGGATCCAGCCAAATCAAGAGAA 960
 DB 901 GGTTTAATATATCAATGAGATACCTTTTGTGGGATCCAGCCAAATCAAGAGAA 960

QY 961 CACATGCGAGTGCAGTGTCTTTCGATGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 DB 961 CACATGCGAGTGCAGTGTCTTTCGATGCTGCTGCTGCTGCTGCTGCTGCTG 1020

QY 1021 AGAGACCGATTAACAAAAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCT 1080
 DB 1021 AGAGACCGATTAACAAAAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCT 1080

QY 1081 GCAGGTGCTGCTGCTTCCCTTAGTGTCTATTTTGAATATACAGTTACATTTG 1140
 DB 1081 GCAGGTGCTGCTGCTTCCCTTAGTGTCTATTTTGAATATACAGTTACATTTG 1140

QY 1141 AGTGGCAGGTTTATTTTATTTGTTGGTATCTGGGTATCAAAAATACATTTCAAT 1200
 DB 1141 AGTGGCAGGTTTATTTTATTTGTTGGTATCTGGGTATCAAAAATACATTTCAAT 1200

QY 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTTTGTGATCTACAT 1260
 DB 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTCTTTTGTGATCTACAT 1260

QY 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGCAATCAAAAATATCAACGATGAAGA 1320
 DB 1261 CTTGTATGTACCTTCCAGCAGGCTTTGGTCTGCAATCAAAAATATCAACGATGAAGA 1320

QY 1321 GTATTTGTTGCTCTATATGCAATCAGTGTCTCTACTTTGCTGGAGTATGTTGCGACTG 1380
 DB 1321 GTATTTGTTGCTCTATATGCAATCAGTGTCTCTACTTTGCTGGAGTATGTTGCGACTG 1380

QY 1381 ATGTTTGAATTTGACTTCCAGTGTCTGATGCTGCTGCAATTTGCTTTTCAAAATGTTTT 1440
 DB 1381 ATGTTTGAATTTGACTTCCAGTGTCTGATGCTGCTGCAATTTGCTTTTCAAAATGTTTT 1440

QY 1441 GAGCACTATTTGGGGATGACATGAAAAGGAAAAATCACTCTGGAGGACAGCTGAT 1500
 DB 1441 GAGCACTATTTGGGGATGACATGAAAAGGAAAAATCACTCTGGAGGACAGCTGAT 1500

QY 1501 GAGGATGACAAAAGAAAACCAAGGAAAATTTGTATGATAAGCAGGTAAGTGAGGAAACAT 1560
 DB 1501 GAGGATGACAAAAGAAAACCAAGGAAAATTTGTATGATAAGCAGGTAAGTGAGGAAACAT 1560

QY 1561 GCMACTGACAGAAAACCACTGAGAGGATTAAGCCCTTAATAAAAAGCATGTCACC 1620
 DB 1561 GCMACTGACAGAAAACCACTGAGAGGATTAAGCCCTTAATAAAAAGCATGTCACC 1620

QY 1621 ATGTTGATGCTGATCTATTTGATGATGTTTGTCTGCTGCACTGTACCTGGGTCAACAGCAAT 1680
 DB 1621 ATGTTGATGCTGATCTATTTGATGATGTTTGTCTGCTGCACTGTACCTGGGTCAACAGCAAT 1680

QY 1681 GCTACTCTAGTCCAGGTGATGCTGCTGCTCATACATCATGATGACGACAGGATATC 1740
 DB 1681 GCTACTCTAGTCCAGGTGATGCTGCTGCTCATACATCATGATGACGACAGGATATC 1740

QY 1741 TTAGATGATTTTGAAGAAGCTTACTTTTGGCTAAGGCAAAAATACAGATGAACATGCACGA 1800
 DB 1741 TTAGATGATTTTGAAGAAGCTTACTTTTGGCTAAGGCAAAAATACAGATGAACATGCACGA 1800

QY 1801 GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTTG 1860
 DB 1801 GTAAATGCTTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTTG 1860

QY 1861 GTGGATTAATAACCTGGAATAACAGCCACATAGCAGTGTGGGAAAAGCTATGCTCTCT 1920
 DB 1861 GTGGATTAATAACCTGGAATAACAGCCACATAGCAGTGTGGGAAAAGCTATGCTCTCTCT 1920

QY 1921 AATGAAAACAGCAGCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTTAT 1980
 DB 1921 AATGAAAACAGCAGCTATAAATCATGAGGACTCTAGATGTAGATTTATGTTTGGTTTAT 1980

QY 1981 TTTGGAGGGGTTATTTGGCTATTTCTGGTGATGATCAACAAATTTCTCTGGATGGTTAGG 2040
 DB 1981 TTTGGAGGGGTTATTTGGCTATTTCTGGTGATGATCAACAAATTTCTCTGGATGGTTAGG 2040

QY 2041 ATAGCTTGAAGGAGAAATCCCAAGACATTCGGGAAAGTACTATTTTATCCCAACAGGGA 2100
 DB 2041 ATAGCTTGAAGGAGAAATCCCAAGACATTCGGGAAAGTACTATTTTATCCCAACAGGGA 2100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 02:41:23 ; Search time: 16661.4 Seconds
(without alignments)
7041.770 Million cell updates/sec

Title: US-10-028-384-1

Perfect score: 2481

Sequence: 1 atggcgagcctcgcccc.....tatctaagaagactgttta 2481

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 2481 | 100.0 | 2481 | 6 | AX799082 Sequence |
| 2 | 2481 | 100.0 | 2481 | 9 | AY074880 Homo sapi |
| 3 | 1276 | 51.4 | 1664 | 6 | CQ782554 Sequence |
| 4 | 1276 | 51.4 | 1664 | 6 | BD127193 Primer fo |
| 5 | 1276 | 51.4 | 1664 | 9 | AK074587 Homo sapi |
| 6 | 972 | 39.2 | 2546 | 6 | AX099510 Sequence |
| 7 | 972 | 39.2 | 2546 | 6 | BD063986 Secreted |
| 8 | 954 | 38.5 | 2510 | 6 | AX136135 Sequence |
| 9 | 954 | 38.5 | 2510 | 6 | BD123520 Secretory |
| 10 | 954 | 38.5 | 2510 | 9 | AK075380 Homo sapi |
| 11 | 861 | 34.7 | 2284 | 6 | AX882932 Sequence |
| 12 | 861 | 34.7 | 2284 | 6 | BD160013 Primer fo |
| 13 | 861 | 34.7 | 2284 | 9 | AK027789 Homo sapi |
| 14 | 828 | 33.4 | 2503 | 6 | AX017997 Sequence |
| 15 | 828 | 33.4 | 2503 | 6 | BD137303 Human nuc |
| 16 | 739 | 29.8 | 957 | 9 | BC015880 Homo sapi |
| 17 | 521 | 21.0 | 764 | 6 | AX136480 Sequence |
| 18 | 521 | 21.0 | 764 | 6 | BD123720 Secretory |
| 19 | 507 | 20.4 | 507 | 6 | AX886102 Sequence |

| | | | | | |
|----|-----|------|--------|---|---------------------|
| 20 | 507 | 20.4 | 507 | 6 | BD025712 Sequence |
| 21 | 506 | 20.4 | 787 | 6 | AX869456 Sequence |
| 22 | 506 | 20.4 | 787 | 6 | BD149518 Primer fo |
| 23 | 502 | 20.2 | 1040 | 6 | CQ723424 Sequence |
| 24 | 430 | 17.3 | 433 | 6 | AX340470 Sequence |
| 25 | 430 | 16.2 | 558 | 6 | AX079689 Sequence |
| 26 | 351 | 14.1 | 616 | 6 | BD210442 Human gen |
| 27 | 332 | 13.0 | 349 | 6 | BD071886 Secreted |
| 28 | 316 | 12.7 | 170691 | 9 | AC104643 Homo sapi |
| 29 | 250 | 10.1 | 250 | 6 | AX885575 Sequence |
| 30 | 250 | 10.1 | 250 | 6 | BD025185 Sequence |
| 31 | 236 | 9.5 | 245 | 6 | AX905227 Sequence |
| 32 | 236 | 9.5 | 245 | 6 | BD040760 Sequence |
| 33 | 236 | 9.5 | 247 | 6 | BD075907 5' EST of |
| 34 | 218 | 8.8 | 503 | 6 | CQ779964 Sequence |
| 35 | 218 | 8.8 | 503 | 6 | CQ781897 Sequence |
| 36 | 218 | 8.8 | 503 | 6 | BD124673 Primer fo |
| 37 | 218 | 8.8 | 503 | 6 | BD126606 Sequence |
| 38 | 218 | 8.8 | 177253 | 9 | AC092024 Homo sapi |
| 39 | 188 | 7.6 | 388 | 6 | AX071612 Sequence |
| 40 | 184 | 7.4 | 737 | 6 | BD020608 Novel gen |
| 41 | 184 | 7.4 | 737 | 6 | BD100546 Novel gen |
| 42 | 179 | 7.2 | 474 | 6 | AX079710 Sequence |
| 43 | 170 | 6.9 | 835 | 9 | HS3424805 Homo sapi |
| 44 | 124 | 5.0 | 355 | 6 | AX340785 Sequence |
| 45 | 114 | 4.6 | 567 | 6 | CQ780783 Sequence |

ALIGNMENTS

| | | | | | |
|------------|--|---------|------|--------|-----------------|
| RESULT 1 | AX799082 | 2481 bp | mRNA | linear | PAT 08-OCT-2003 |
| LOCUS | Sequence 1 from Patent WO03054008. | | | | |
| DEFINITION | AX799082 | | | | |
| ACCESSION | AX799082.1 | | | | |
| VERSION | AX799082.1 | | | | |
| KEYWORDS | Homo sapiens (human) | | | | |
| SOURCE | Homo sapiens | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| REFERENCE | Perreault, C. and Mcbride, K. | | | | |
| AUTHORS | Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy | | | | |
| TITLE | Patent: WO 03054008-A 1 03-JUL-2003; | | | | |
| JOURNAL | Comptatigene Inc. (CA) | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..2481 | | | | |
| | /organism="Homo sapiens" | | | | |
| | /mol_type="mRNA" | | | | |
| | /db_xref="taxon:9606" | | | | |
| | 1..2481 | | | | |
| | /codon_start=1 | | | | |
| | /product="SIMP" | | | | |
| | /protein_id="CAE48447.1" | | | | |
| | /db_xref="GI:37605058" | | | | |
| | /translation="MAEPSAPESKHSKSLNSGFWGLMALGNSRHGHGPGQAQCAHKA AGAAPPKAPAGLSGSLQSPAGWQSLISFTILFLAWLAGFSRLFAVIRPESITHEF DPWFYVSTHLLASHGVEYELNWFDRVAVPLGRIVGVTVPLGLMTAGLHWILNTL NITHIDVCVFLAPTFSGLTSLSTELLRELWNQAGLLAACFAIVGVYTSRVAG SFDEGTAIFALQTYIILWKSIVKSGVFWMCCLLSYFQFIRVSEHNAAGVAFALLQAY FVLLLMORVSKRYIAYSTFYIVGLILSNQIPFVGFQFIRVSEHNAAGVAFALLQAY AFLQYLRDLTKQEFQTLFELGSLAAGAVFLSVILTYITGTPAGWGRFVSLWDGTG AKIHIPIASVSEHQPTTWYFFDFLHLVCTFPAFLWFCIKINDEKRVFVALYISA VYFAGVMVRLMLTIPVVCMLSAIAFSPVFEHLGDMKRENPVPTVSDSDDEKNOQ NLYDKAGKVRKHATIQEKEGLGPNKISIVTLMMLMFMFAVCTVTSNAYSQPS VTNASHYALNGKAMSNSETAAVKIMETLDVDVVLVIFGVGIGSSGDDINKLTMWVIA EGEHPKDIRSDYETFGEDVQKAGSTILNCLMYKMSYVFGQMQLDPTPGPDR TRNIEGNKDKIKHLEAETSEHLVRIYKVKAPDNRETLDHKEKRVNIFPKQKLS KKTTRKRKRGYIKNKLVPFKKKSKKTV" | | | | |

CDS

~~Dupli~~

Blank Page

| | | | |
|----|------|--|------|
| Db | 458 | GATATCAACAATAATTCCTCGATGGTTTAGGATAGCTGAAGGAGAACATCCCAAAGACATT | 517 |
| Qy | 2071 | CGGAAAGTGACTATTTTCCACACAGGAGAAATCCGTTGACACAAAGACAGGATCCGCT | 2130 |
| Db | 518 | CGGAAAGTGACTATTTTCCACACAGGAGAAATCCGTTGACACAAAGACAGGATCCGCT | 577 |
| Qy | 2131 | ACTTTGTTGAAATTCGCTTATGTATATAAAATGTGCATCTACAGATTTGGAGAAATCAGCTG | 2190 |
| Db | 578 | ACTTTGKTGAATTCGCTTATGTATATAAAATGTGCATCTACAGATTTGGAGAAATCAGCTG | 637 |
| Qy | 2191 | GATTTTCGTACACCCCGAGGTTTGGACCGAACACGTAAATCTGAGATTTGGAATAAAGGAC | 2250 |
| Db | 638 | GATTTTCGTACACCCCGAGGTTTGGACCGAACACGTAAATCTGAGATTTGGAATAAAGGAC | 697 |
| Qy | 2251 | ATTAATAATCAACATTTTGGAGAGAGCCTTTACATCAGACACACTGGCTGTGTAGATATAT | 2310 |
| Db | 698 | ATTAATAATCAACATTTTGGAGAGAGCCTTTACATCAGACACACTGGCTGTGTAGATATAT | 757 |
| Qy | 2311 | AAAGTAAAGCACCTGATTAACAGGAGAGACATTAGATTCACAAACCTCGAGTCACCAACATT | 2370 |
| Db | 758 | AAAGTAAAGCACCTGATTAACAGGAGAGACATTAGATTCACAAACCTCGAGTCACCAACATT | 817 |
| Qy | 2371 | TTCCCAAAACAGAAAGTATTTGTCAAAAGAGACTTACAAAGGAAGCGTGGCTACATTAAA | 2430 |
| Db | 818 | TTCCCAAAACAGAAAGTATTTGTCAAAAGAGACTTACAAAGGAAGCGTGGCTACATTAAA | 877 |
| Qy | 2431 | AATAAGCTGGTTTTTAAGAAAGCAGAAAAATATCTAAGAGACGCTGTTAA | 2481 |
| Db | 878 | AATAAGCTGGTTTTTAAGAAAGCAGAAAAATATCTAAGAGACGCTGTTAA | 2928 |

Search completed: December 15, 2004, 15:46:57
Job time : 1672 secs

```

Db 543 TCATCTACAGATTGGGAGAAATGAGTGGATTTTGTACACCCAGGTTTGACCGA 602
QY 2221 ACACGTAAATGCTGAGATTGGAAATAGGACATTAAATTTCAACATTGGGAAGGCTTT 2280
Db 603 ACACGTAAATGCTGAGATTGGAAATAGGACATTAAATTTCAACATTGGGAAGGCTTT 662
QY 2281 ACATCAGAACACTGGCTGTTAGGATATATAGATTAAGATTAAGACACTGATACAGGGAGCA 2340
Db 663 ACATCAGAACACTGGCTGTTAGGATATATAGATTAAGATTAAGACACTGATACAGGGAGCA 722
QY 2341 TTAGATCAAAACCTCGAGTCACCAACATTTTCCAAAACAGAAAGTATTGTCAAGAAG 2400
Db 723 TTAGATCAAAACCTCGAGTCACCAACATTTTCCAAAACAGAAAGTATTGTCAAGAAG 782
QY 2401 ACTACCAAAGGAGGCTGGGTACATTAAATAAGCTGTTTAAAGAAAGGCAAGAA 2460
Db 793 ACTACCAAAGGAGGCTGGGTACATTAAATAAGCTGTTTAAAGAAAGGCAAGAA 842
QY 2461 ATAATCAAGAGACTGTTTAA 2481
Db 843 ATATCTAAGAGACTGTTTAA 863

RESULT 15
AAD08315
ID AAD08315 standard; cDNA; 2547 BP.
AC
AC
AD08315;
DT 08-AUG-2001 (first entry)
DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 43.
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vunerary; binding partner identification;
KW gene therapy; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 68..928
XX /*tag= a
XX /product= "Human secreted protein precursor"
XX /trans_except= (pos:584..586, aa:Xaa)
XX /note= "Xaa equals any of the naturally occurring L-amino
XX acids"
XX sig_peptide 68..130
XX /*tag= b
XX mat_peptide 131..925
XX /*tag= c
XX /product= "Mature human secreted protein"
XX
XX WO200136440-A1.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US031282.
XX
XX 19-NOV-1999; 99US-0166414P.
XX 21-JUL-2000; 2000US-0219665P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Birse CE, Moore PA;
XX

```

```

DR WPI; 2001-343795/36.
DR P-ESDB; AAE03850.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Claim 1; Page 462-463; 553pp; English.
XX
XX AAD08283-AA08355 represent cDNAs corresponding to 23 human secreted
XX protein genes, and AAE03818-AA03870 represent the proteins they encode.
XX AAE03871-AA03896 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 23 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infections. The
XX proteins can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
XX present sequence represents a human secreted protein-encoding cDNA of the
XX invention
XX
XX SQ Sequence 2547 BP; 833 A; 422 C; 492 G; 798 T; 0 U; 2 Other;

```

```

Query Match 33.9%; Score 840; DB 4; Length 2547;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 890; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1591 TTAGGCCCTAATAAAGAGCATTTGTCACCATGTTTGATGCTGATGCTATTGATGATGTTT 1650
Db 38 TTAGGCCCTAATAAAGAGCATTTGTCACCATGTTTGATGCTGATGCTATTGATGATGTTT 97
QY 1651 GCTGTCCACTGTACTGGGTCCACAGCAATGCTTACTCTAGTCCAGGTAGTCTGCGCC 1710
Db 98 GCTGTCCACTGTACTGGGTCCACAGCAATGCTTACTCTAGTCCAGGTAGTCTGCGCC 157
QY 1711 TCATACAATCATGATGGCCACCAGGAATATCTTTAGATGATTTTAGAGAAGCTTACTTTGG 1770
Db 158 TCATACAATCATGATGGCCACCAGGAATATCTTTAGATGATTTTAGAGAAGCTTACTTTGG 217
QY 1771 CTAAGGCAAAATACAGATGAACATGCACGAGTAATGCTTGTGGGATATGGGTATCAG 1830
Db 218 CTAAGGCAAAATACAGATGAACATGCACGAGTAATGCTTGTGGGATATGGGTATCAG 277
QY 1831 ATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAAATACACCTGGAATAACAGCCAC 1890
Db 278 ATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAAATACACCTGGAATAACAGCCAC 337
QY 1891 ATAGCACTGGTGGAAAGCTATGCTTCTAATGAACAGCAGCCTATAAATCATGAGG 1950
Db 338 ATAGCACTGGTGGAAAGCTATGCTTCTAATGAACAGCAGCCTATAAATCATGAGG 397
QY 1951 ACTCTAGATGTAGATTATGTTTGGTATTTTGGAGGGGTTTATGGCTATTCTGGTGAT 2010
Db 398 ACTCTAGATGTAGATTATGTTTGGTATTTTGGAGGGGTTTATGGCTATTCTGGTGAT 457
QY 2011 GRTATCAACAAATTTCTCTGGATGTTAGGTAGCTGAAGGAGAACATCCCCAAGACATT 2070

```


Db 421 AGGACTCTAGATGTAGATTAAGTTTGGTTATTTTGGAGGGGTTATTGGCTATTCTGGT 480

Qy 2008 GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTCAAGGAGAAATCCCAAGAC 2067

Db 481 GATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTCAAGGAGAAATCCCAAGAC 540

Qy 2068 ATTCGGGAAAGTGACTATTTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGATCC 2127

Db 541 ATTCGGGAAAGTGACTATTTTACCCACAGGAGAAATTCGGTGTAGACAAAGCAGATCC 600

Qy 2128 CCTACTTTGTTGAATTCCTTATGTATATAAATCTCATCTACAGATTTGGAGAAATGCAG 2187

Db 601 CCTACTTTGTTGAATTCCTTATGTATATAAATCTCATCTACAGATTTGGAGAAATGCAG 660

Qy 2188 CTGGATTTCTGTACACCCCGAGTTTTCACCGACACGTAATGCTGAGATTTGGAATTAAG 2247

Db 661 CTGGATTTCTGTACACCCCGAGTTTTCACCGACACGTAATGCTGAGATTTGGAATTAAG 720

Qy 2248 GACATTAATTTCAACATTTTGAAGAGCCCTTTACATCAGAACACTGGCTTTAGGATA 2307

Db 721 GACATTAATTTCAACATTTTGAAGAGCCCTTTACATCAGAACACTGGCTTTAGGATA 780

Qy 2308 TATTAAGTAAAGCACTGTATACAGGAGACATTAGATCAAAACCTCGAGTCAACAC 2367

Db 781 TATTAAGTAAAGCACTGTATACAGGAGACATTAGATCAAAACCTCGAGTCAACAC 840

Qy 2368 ATTTTCCCAAAACAGAAAGTATTTGTCAAGAGACCTACCAAGGAAGCGTGGTACATT 2427

Db 841 ATTTTCCCAAAACAGAAAGTATTTGTCAAGAGACCTACCAAGGAAGCGTGGTACATT 900

Qy 2428 AATAATAGCTGGTTTATAGAGGACGAGAAATATCTAAGAACACTGTTTAA 2481

Db 901 AATAATAGCTGGTTTATAGAGGACGAGAAATATCTAAGAACACTGTTTAA 954

RESULT 13

AA08289

ID AA08289 standard; cDNA; 2537 BP.

XX

AC AA08289;

XX

DT 08-AUG-2001 (first entry)

XX

DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 17.

XX

KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma; fetal abnormality; developmental abnormality; haematopoietic disorder; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; Parkinson's disease; cognitive disorder; schizophrenia; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder; inflammation; neurological disorder; Alzheimer's disease; food additive; angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy; pregnancy-related disorder; endocrine disorder; infection; wound healing; cell culture; chemotaxis; vulnery; binding partner identification;

XX

OS Homo sapiens.

XX

FH Location/Qualifiers

FT CDS 93..953

FT sig_peptide /tag= a

FT /product= "Human secreted protein precursor"

FT mat_peptide /tag= b

FT /tag= c

XX

FN WO200136440-A1.

XX

PD 25-MAY-2001.

XX

PF 15-NOV-2000; 2000WO-US031282.

XX

PR 19-NOV-1999; 99US-0166414P.

PR 21-JUL-2000; 2000US-0219665P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

FI Ruben SM, Komatsoulis GA, Birse CE, Moore PA;

FI WPI; 2001-343795/36.

DR P-PSDB; AA03824.

XX

PT Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition.

XX

PS Claim 1; Page 443; 553pp; English.

XX

CC AAD08283-AD08355 represent cDNAs corresponding to 23 human secreted protein genes, and AA03818-AA03870 represent the proteins they encode. CC AA03871-AA03896 represent human secreted protein fragments or variants. CC The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. CC Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 23 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The CC present sequence represents a human secreted protein-encoding cDNA of the CC invention

XX

SQ Sequence 2537 BP; 816 A; 434 C; 487 G; 800 T; 0 U; 0 Other;

Query Match 35.9%; Score 891; DB 4; Length 2537;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1591 TTAGGCCCTTAATATAAAGCAATTCACCATGTTGATGCTGATGCTATTGATGATGTTT 1650

Db 63 TTAGGCCCTTAATATAAAGCAATTCACCATGTTGATGCTGATGCTATTGATGATGTTT 122

Qy 1651 GCTGTCCACTGTACTCTGGGTACAAGCAATGCTACTTACTTCAAGTAGTCTCTGCC 1710

Db 123 GCTGTCCACTGTACTCTGGGTACAAGCAATGCTACTTACTTCAAGTAGTCTCTGCC 182

Qy 1711 TCATACAATCATGTGGGACAGGCAATATCTTAGATGATTTTACAGAAAGCTTACTTTGG 1770

Db 183 TCATACAATCATGTGGGACAGGCAATATCTTAGATGATTTTACAGAAAGCTTACTTTGG 242

Qy 1771 CTAAGGCAAAATACAGATGAACATGACAGTAATATGCTTGGTGGGATTTATGCTATCAG 1830

Db 243 CTAAGGCAAAATACAGATGAACATGACAGTAATATGCTTGGTGGGATTTATGCTATCAG 302

Qy 1831 ATAGCTGGAAATCGCTAATAGAACTACGTTGGTGGATATACACCTCGAATACAGCCAC 1890

Db 303 ATAGCTGGAAATCGCTAATAGAACTACGTTGGTGGATATACACCTCGAATACAGCCAC 362

Qy 1891 ATAGCACTGGTGGGAAAGCAATATGCTTCTTAATGAACAGCAGCCTATAAATCATGAGG 1950

QY 1810 TGGTGGGATTATGGCTATCAGATAGCTGGATGGCTAATAGAACTACCTTGGTGGATAAT 1869
 Db 301 TGGTGGGATTATGGCTATCAGATAGCTGGATGGCTAATAGAACTACCTTGGTGGATAAT 360
 QY 1870 AACACCTGGGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTTCTAATGAACA 1929
 Db 361 AACACCTGGGAATAACAGCCACATAGCACTGGTGGGAAAGCTATGCTCTTCTAATGAACA 420
 QY 1930 GCACCTATTAATCATGAGGCTCTAGATGTAGATTTGTTGTTATTTGGTGGAGG 1989
 Db 421 GCACCTATTAATCATGAGGCTCTAGATGTAGATTTGTTGTTATTTGGTGGAGG 480
 QY 1990 GTTATTGGCTATTCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 2049
 Db 481 GTTATTGGCTATTCTGGTGTATGATATCAACAAATTTCTCTGGATGGTTAGGATAGCTGAA 540
 QY 2050 GGAGACATCCCAAGACATTCGGGAAGTGTATTTTACCCACAGGAGNATCCGT 2109
 Db 541 GGAGACATCCCAAGACATTCGGGAAGTGTATTTTACCCACAGGAGNATCCGT 600
 QY 2110 GTACAAAAGCAGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCATCTAC 2169
 Db 601 GTACAAAAGCAGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCATCTAC 660
 QY 2170 AGATTGGGAATGCGAGTGGATTTTCTGTACACCCCGAGTTTGGCCGACAGGTAA 2229
 Db 661 AGATTGGGAATGCGAGTGGATTTTCTGTACACCCCGAGTTTGGCCGACAGGTAA 720
 QY 2230 GCTGAGATTGGAATAAGGACATTAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 2289
 Db 721 GCTGAGATTGGAATAAGGACATTAATCAACATTTTGGAAAGAGCTTTTACATCAGAA 780
 QY 2290 CACTGGCTTGTAGGATATATAAGTAAGCAAGCACTGTAAACAGGAGACATAGATCAC 2349
 Db 781 CACTGGCTTGTAGGATATATAAGTAAGCAAGCACTGTAAACAGGAGACATAGATCAC 840
 QY 2350 AAACCTCGAGTCAACCAATTTCCCAAAACAGAGTATTTGTCAAAAGAGCTACCAAA 2409
 Db 841 AAACCTCGAGTCAACCAATTTCCCAAAACAGAGTATTTGTCAAAAGAGCTACCAAA 900
 QY 2410 AGGAAGCTGGCTACATTAATAAATAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAG 2469
 Db 901 AGGAAGCTGGCTACATTAATAAATAGCTGGTTTAAAGAAAGGCAAGAAATATCTAAG 960
 QY 2470 AAGACTGTTAA 2481
 Db 961 AAGACTGTTAA 972

RESULT 12
 AAF93772

ID AAF93772 standard; cDNA; 2510 BP.

XX AC AAF93772;

XX DT 23-MAY-2001 (first entry)

XX DE Human cDNA encoding a membrane or secretory protein clone: PSE0070.

XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;

XX KW rheumatoid arthritis; diabetes; ss.

XX OS Homo sapiens.

XX PN EPI067182-A2.

XX PD 10-JAN-2001.

XX PF 07-JUL-2000; 2000EP-00114090.

XX PR 08-JUL-1999; 99JP-00194179.

XX PR 11-JAN-2000; 2000JP-00118775.

XX PR 02-MAY-2000; 2000JP-00183766.

XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPI; 2001-093989/11.
 XX P-PSDB; AAB88345.
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
 XX gene therapy or as candidate target molecules in drug development.
 XX Claim 1; SEQ ID NO 57; 609pp + Sequence Listing; English.
 XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
 XX which encode human secretory or membrane proteins represented by AAB88317
 XX - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 XX AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
 XX invention. The invention also includes methods for the production of
 XX antibodies directed against the proteins, and cDNA sequences, which can
 XX be used in vaccines. The polynucleotide sequences can be used in gene
 XX therapy. The polynucleotide sequences and the proteins they encode may be
 XX used in the prevention, treatment and diagnosis of diseases associated
 XX with inappropriate secretory protein/membrane protein expression. The
 XX nucleic acids and complementary sequences may also be used as DNA probes
 XX in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 XX and quantitate the presence of similar nucleic acid sequences in samples.
 XX They may also be used to study the expression and function of secretory
 XX proteins/membrane polypeptides and their role in metabolism. The
 XX polypeptides may be used as antigens in the production of antibodies
 XX against them and in assays to identify modulators (agonists and
 XX antagonists) of expression and activity. The antibodies and antagonists
 XX may also be used as therapeutic agents to down regulate expression and
 XX activity. The antibodies may also be used as diagnostic agents for
 XX detecting the presence of the polypeptides in samples (e.g. by enzyme
 XX linked immunosorbent assay (ELISA)). Examples of diseases which may be
 XX treated include rheumatoid arthritis and diabetes

XX Sequence: 2510 BP; 802 A; 412 C; 489 G; 807 T; 0 U; 0 Other;

Query Match 38.5%; Score 954; DB 5; Length 2510;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1528 TTGTTATGATTAAGGAGGTAAGTGAAGGAAACATGCAACAGGAAAAAAGCTGAAGAG 1587
 Db 1 TTGTTATGATTAAGGAGGTAAGTGAAGGAAACATGCAACAGGAAAAAAGCTGAAGAG 60
 QY 1588 GGATTAGGCCCTTAATATAAAAAGCATTTGTCCACATGTTGATGCTGATGCTATTGATGATG 1647
 Db 61 GGATTAGGCCCTTAATATAAAAAGCATTTGTCCACATGTTGATGCTGATGCTATTGATGATG 120
 QY 1648 TTGCTGTCCACTGTACTGGCTCACAAGCAATGCCCTACTCTAGTCCAAAGTGTAGTCCCTG 1707
 Db 121 TTGCTGTCCACTGTACTGGCTCACAAGCAATGCCCTACTCTAGTCCAAAGTGTAGTCCCTG 180
 QY 1708 GCTTCATACATCATGATGGCCACCAGCAATATCTTAGATGATTTTAGAGAAGCTTACTTT 1767
 Db 181 GCTTCATACATCATGATGGCCACCAGCAATATCTTAGATGATTTTAGAGAAGCTTACTTT 240
 QY 1768 TGGCTAAGGCAAAATACAGATGAACATGACGAGTATGCTCTGGTGGGATATGCTAT 1827
 Db 241 TGGCTAAGGCAAAATACAGATGAACATGACGAGTATGCTCTGGTGGGATATGCTAT 300
 QY 1828 CAGATAGCTGGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGAATAACAGC 1887
 Db 301 CAGATAGCTGGAATGGCTTAATAGAACTACGTTGGTGGATAATAACACCTGGAATAACAGC 360
 QY 1888 CACATAGCACTGGTGGGAAAGCTATGCTCTTCTTAAGTAACACGACGACCTATAAATCATG 1947
 Db 361 CACATAGCACTGGTGGGAAAGCTATGCTCTTCTTAAGTAACACGACGACCTATAAATCATG 420
 QY 1948 AGGACTCTAGATGTAGATTAATGTTTGGTATTATTTTGGAGGGTTATTGGCTATTCTGTT 2007

| | | | |
|-----------|---|---|------|
| QY | 1227 | GACTGGGTCCTTCCTTCCTGATCTACATATCTTCCTGATGTAACCTTCCAGCAGCGCT | 1286 |
| DB | 961 | GACTTGGGTCCTTCCTTCCTGATCTACATATCTTCCTGATGTAACCTTCCAGCAGCGCT | 1020 |
| QY | 1287 | TTGGTTCCTGCATCAAAAATATCAACGATCAAGA | 1320 |
| DB | 1021 | TTGGTTCCTGCATCAAAAATATCAACGATCAAGA | 1054 |
| RESULT 10 | | | |
| ID | AAV44866 | | |
| XX | AAV44866 | standard; CDNA; 2546 BP. | |
| XX | AAV44866; | | |
| XX | 21-OCT-1998 | (first entry) | |
| XX | Clone CT585_1 | coding sequence. | |
| XX | Secreted protein; nutritional source; cell proliferation activity; | | |
| KW | cell differentiation activity; immune stimulant; tissue growth activator; | | |
| KW | haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor; | | |
| KW | tumour inhibitor; clone CT585_1; ds. | | |
| XX | Homio sapiens. | | |
| XX | Key | Location/Qualifiers | |
| FF | CDS | 112..972 | |
| FT | | /*tag= a | |
| XX | WO9825962-A2. | | |
| PN | 18-JUN-1998. | | |
| PD | 12-DEC-1997; | 97WO-US023224. | |
| PF | 13-DEC-1996; | 96US-00766263. | |
| PR | 11-DEC-1997; | 97US-00989232. | |
| XX | (GENY) GENETICS INST INC. | | |
| XX | Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M; | | |
| PI | Spaulding V, Agostino MU; | | |
| XX | WPI; 1998-362424/31. | | |
| DR | P-PSDB; AAW69247. | | |
| DR | New isolated polynucleotides - obtained from human adult testis, human | | |
| PT | adult ovary, human adult brain and human adult heart CDNA libraries. | | |
| XX | Claim 35; Page 79-81; 108pp; English. | | |
| XX | This sequence represents a polynucleotide of the invention, and encodes a | | |
| XX | secreted protein. It was isolated from a human adult brain CDNA library, | | |
| CC | and is designated clone CT585_1. The DNA sequences and encoded | | |
| CC | polypeptides can be used as nutritional sources or supplements, or may | | |
| CC | exhibit e.g. cytokine and cell proliferation/differentiation activity, | | |
| CC | immune stimulating or suppressing activity, haematopoiesis regulating | | |
| CC | activity, receptor/ligand activity, anti-inflammatory activity, | | |
| CC | activin/inhibin activity, chemostatic/chemokinetic activity, | | |
| CC | cadherin/tumour invasion suppressor activity, tissue growth activity, | | |
| CC | tumour inhibition activity or other activities | | |
| XX | Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other; | | |

```

Query Match      39.3%; Score 972; DB 2; Length 2546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1510 AAAAGAAACCAAGGAAATTTGTATGATAGGCAGGTAAAGTGCAGGAAACATGCAACTGAA 1569

Db 1 AAAAGAAACCAAGGAAATTTGTATGATAGGCAGGTAAAGTGCAGGAAACATGCAACTGAA 60

```


QY 567 AAGAGAACTTTGGAAACCAAGAGCAGGACTTTTAGCTGCTTTTATTTATGCTATTTGACC 626
 DB 301 AAGAGAACTTTGGAAACCAAGAGCAGGACTTTTAGCTGCTTTTATTTATGCTATTTGACC 360
 QY 627 AGGCTACATATCTCGGTGAGTACGCTGATCTTTGATATGAGGCACTTTGCTATTTTGGC 686
 DB 361 AGGCTACATATCTCGGTGAGTACGCTGATCTTTGATATGAGGCACTTTGCTATTTTGGC 420
 QY 687 ACTTCAGTTCACATATCTTTGATATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 746
 DB 421 ACTTCAGTTCACATATCTTTGATATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 480
 QY 747 AATGCTGCTGCTTATCTTATTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 806
 DB 481 AATGCTGCTGCTTATCTTATTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 540
 QY 807 CATCAATCTTATCTTATCTTATTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 866
 DB 541 CATCAATCTTATCTTATCTTATTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 600
 QY 867 AGTCTACATATGATATCTTATTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 926
 DB 601 AGTCTACATATGATATCTTATTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 660
 QY 927 TTTTGTGGGATTCAGGCACTTATCTTATTTATGAGGCACTTTGATATGAGGCACTTTTGGC 986
 DB 661 TTTTGTGGGATTCAGGCACTTATCTTATTTATGAGGCACTTTGATATGAGGCACTTTTGGC 720
 QY 987 ATTGCTGCAAGCTTATGCTTTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 1046
 DB 721 ATTGCTGCAAGCTTATGCTTTTATGAGGCACTTTGATATGAGGCACTTTGCTATTTTGGC 780
 QY 1047 CCAGACCTTTTCTTTTGGGATTCAGGCACTTATCTTATTTATGAGGCACTTTGCTATTTTGGC 1106
 DB 781 CCAGACCTTTTCTTTTGGGATTCAGGCACTTATCTTATTTATGAGGCACTTTGCTATTTTGGC 840
 QY 1107 CTATTTGCACTTATACAGGTTACATTTGAGGCACTTATCTTATTTATGAGGCACTTTTGGC 1166
 DB 841 CTATTTGCACTTATACAGGTTACATTTGAGGCACTTATCTTATTTATGAGGCACTTTTGGC 900
 QY 1167 TACTGGGTATCAAAATACATCTTATTTATGAGGCACTTATCTTATTTATGAGGCACTTTTGGC 1226
 DB 901 TACTGGGTATCAAAATACATCTTATTTATGAGGCACTTATCTTATTTATGAGGCACTTTTGGC 960
 QY 1227 GACTTGGGTGCTTTCTTTTGGGATTCAGGCACTTATCTTATTTATGAGGCACTTTTGGC 1286
 DB 961 GACTTGGGTGCTTTCTTTTGGGATTCAGGCACTTATCTTATTTATGAGGCACTTTTGGC 1020
 QY 1287 TTGGTTCGATCAAAATACATCTTATTTATGAGGCACTTATCTTATTTATGAGGCACTTTTGGC 1320
 DB 1021 TTGGTTCGATCAAAATACATCTTATTTATGAGGCACTTATCTTATTTATGAGGCACTTTTGGC 1054

RESULT 8

ADG78372

ID ADG78372 standard; cDNA; 1543 BP.

XX AC

XX AC ADG78372;

XX DT 11-MAR-2004 (first entry)

XX DE Human secreted protein cDNA #123.

XX KW

Human; secreted protein; gene; ss; immune disorder;
 haematopoietic disorder; hyperproliferative disorder; infectious disease;
 inflammatory disorder; HIV; anaemia; thrombocytopenia; bleeding; stroke;
 myocardial infarction; Addison's disease; rheumatoid arthritis;
 dermatitis; Grave's disease; multiple sclerosis; glomerulonephritis;
 diabetes; graft-versus-host disease; inflammatory bowel disease; cancer;
 bacterial infection; viral infection; parasitic infection; osteoporosis;
 wound; atherosclerosis; Alzheimer's disease; Parkinson's disease; ulcer;
 food additive; food preservative.

XX

OS Homo sapiens.
 XX US2003211472-A1.
 XX 13-NOV-2003.
 XX 28-MAR-2001; 2001US-00818683.
 XX 05-MAY-1999; 99US-00305736.
 XX (FENG/) FENG P.
 XX (RUBEN/) RUBEN S M.
 XX (ROSE/) ROSEN C A.
 XX (EBNER/) EBNER R.
 XX (OLSEN/) OLSEN H S.
 XX (NIJ/) NI J.
 XX (WEI/) WEI Y.
 XX (SOPP/) SOPPET D R.
 XX (MOORE/) MOORE P A.
 XX (KYAW/) KYAW H.
 XX (LAFLEUR/) LAFLEUR D W.
 XX (SHI/) SHI Y.
 XX (JANAT/) JANAT F.
 XX (ENDRESS/) ENDRESS G A.
 XX (CARTER/) CARTER K C.
 XX Feng P, Ruben SM, Rosen CA, Ebner R, Olsen HS, Ni J, Wei Y;
 PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
 PI Endress GA, Carter KC;
 XX WPI: 2004-051575/05.
 XX P-PSDB; ADG78500.

XX New secreted nucleic acid for diagnosing, preventing or treating diseases
 PT associated with aberrant expression or activity of the polypeptide it
 PT encodes, e.g. cancer, human immunodeficiency virus, Parkinson's disease,
 PT or diabetes.

XX Claim 1; SEQ ID NO 133; 377pp; English.

XX The invention also relates to human secreted proteins and the nucleic
 CC acids encoding them. The proteins and nucleic acids are useful in
 CC diagnosing, preventing, prognosing or treating diseases or disorders
 CC associated with aberrant expression and/or activity of the secreted
 CC proteins, such as immune disorders, haematopoietic disorders,
 CC hyperproliferative disorders, infectious diseases or inflammatory
 CC disorders. In particular, the diseases or disorders are HIV, anaemia,
 CC thrombocytopenia, bleeding, stroke, myocardial infarction, Addison's
 CC disease, rheumatoid arthritis, dermatitis, Grave's disease, multiple
 CC sclerosis, glomerulonephritis, diabetes, graft-versus-host disease,
 CC inflammatory bowel disease, cancer, bacterial infections, viral
 CC infections, parasitic infections, osteoporosis, wounds, atherosclerosis,
 CC Alzheimer's disease, Parkinson's disease or ulcers. The sequences may
 CC also be used as food additives or preservatives, or for modulating
 CC mammalian mental or physical characteristics. The nucleic acids are also
 CC used in chromosome mapping, in forensic biology or as molecular weight
 CC markers. This sequence represents cDNA encoding a human secreted protein
 CC of the invention.

XX Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Query Match 42.5%; Score 1054; DB 12; Length 1543;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CTTCCCGCTATCCGCTTCGAAAGCATCATCCAGAGTTCGACCGCTGGTTAACTATAG 326
 DB 1 CTTCCCGCTATCCGCTTCGAAAGCATCATCCAGAGTTCGACCGCTGGTTAACTATAG 60
 QY 327 ATCAACACATCATCTTGATCTCATGGTTCATGAATTTTAAATGTTGATGAAG 386
 DB 61 ATCAACACATCATCTTGATCTCATGGTTCATGAATTTTAAATGTTGATGAAG 120

```
Db 541 CATCACTTATTCACATGCAATGATTTGTTGTTACTGATGACAGATACAGCAAG 600
QY 867 AGTCTACATAGCATATAGCACTTTCTACATGTTGGGTTTAATTAATCAATGAGATACC 926
Db 601 AGTCTACATAGCATATAGCACTTTCTACATGTTGGGTTTAATTAATCAATGAGATACC 660
QY 927 TTTTGTGGGATCCAGCAATCAGAAACAAAGTGAACACATGGCAGCTGCGAGGTCTTTTGC 986
Db 661 TTTTGTGGGATCCAGCAATCAGAAACAAAGTGAACACATGGCAGCTGCGAGGTCTTTTGC 720
QY 987 ATTGCTGAAGCTTATGCTTTCTTTCGAGTATCTGAGAGACCGATTAAACAAACAGAGTT 1046
Db 721 ATTGCTGAAGCTTATGCTTTCTTTCGAGTATCTGAGAGACCGATTAAACAAACAGAGTT 780
QY 1047 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTGCGAGTGTCTTCTTCTAGTGTCTAT 1106
Db 781 CCAGACCCCTTTCTTTTGGGTGATCACTAGCTGCGAGTGTCTTCTTCTAGTGTCTAT 840
QY 1107 CTATTTGACTTATACAGTTACATGTCACCATGAGTGGCAGGTTTATTCATTTGGGA 1166
Db 841 CTATTTGACTTATACAGTTACATGTCACCATGAGTGGCAGGTTTATTCATTTGGGA 900
QY 1167 TACTGGGTATGCAAAATACACATTCATTCATGATGATGTCGAGCATCAACCTTAC 1226
Db 901 TACTGGGTATGCAAAATACACATTCATTCATGATGATGTCGAGCATCAACCTTAC 960
QY 1227 GACTGGGTGCTTCTTCTTGTGATCTACATATTTCTGTATGATGATGTCGAGCAGGCT 1286
Db 961 GACTGGGTGCTTCTTCTTGTGATCTACATATTTCTGTATGATGATGTCGAGCAGGCT 1020
QY 1287 TTGGTTCTGCTCAAAATATCAACGATGAAAGA 1320
Db 1021 TTGGTTCTGCTCAAAATATCAACGATGAAAGA 1054
```

RESULT 7

```
ACD18981
ID ACD18981 standard; cDNA; 1543 BP.
XX AC ACD18981;
XX DT 21-AUG-2003 (first entry)
XX DE Novel human secreted protein cDNA #123.
XX KW Human; secreted protein; cytostatic; neuroprotective; hepatotropic;
KW Gene therapy; cancer; liver disorder; hepatitis; neural disorder;
KW Alzheimer's disease; gene; ss.
XX OS Homo sapiens.
XX PN US2003028003-A1.
XX PD 06-FEB-2003.
XX PF 12-OCT-2001; 2001US-00974879.
XX 07-NOV-1997; 97US-0064900P.
PR 07-NOV-1997; 97US-0064908P.
PR 07-NOV-1997; 97US-0064911P.
PR 07-NOV-1997; 97US-0064912P.
PR 07-NOV-1997; 97US-0064983P.
PR 07-NOV-1997; 97US-0064984P.
PR 07-NOV-1997; 97US-0064985P.
PR 07-NOV-1997; 97US-0064987P.
PR 07-NOV-1997; 97US-0064988P.
PR 17-NOV-1997; 97US-0066089P.
PR 17-NOV-1997; 97US-0066090P.
PR 17-NOV-1997; 97US-0066094P.
PR 17-NOV-1997; 97US-0066095P.
PR 17-NOV-1997; 97US-0066100P.
PR 04-NOV-1998; 98WO-US023435.
```

```
PR 05-MAY-1999; 99US-00305736.
PR 13-OCT-2000; 2000US-0239893P.
XX 28-MAR-2001; 2001US-00818683.
PA (ROSE/) ROSEN C A.
PA (FENG/) FENG P.
PA (RUBEN/) RUBEN S M.
PA (EBNER/) EBNER R.
PA (OLSEN/) OLSEN H S.
PA (NIJ/) NI J.
PA (WEI/) WEI Y.
PA (SOPP/) SOPPET D R.
PA (MOORE/) MOORE P A.
PA (KYAW/) KYAW H.
PA (LAFLEUR/) LAFLEUR D W.
PA (SHI/) SHI Y.
PA (JANAT F.) JANAT F.
PA (ENDRESS G A.) ENDRESS G A.
PA (CASTER K C.) CASTER K C.
PA (BIRSE C E.) BIRSE C E.
XX PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;
PI Endress GA, Carter KC, Birse CE;
XX WPI: 2003-479549/45.
XX P-PSDB; ABO14094.
XX The invention describes a new isolated nucleic acid molecule comprising a
XX sequence having at least 95% identity with a sequence comprising: (a) a
XX polynucleotide (PN) fragment of a sequence comprising 420-3435 bp, or its
XX allelic variant; (b) a PN fragment of the cDNA sequence; (c) a PN
XX sequence encoding a polypeptide, or its fragment, domain, epitope or
XX species homologue; or (d) a PN that hybridises under stringent conditions
XX to any one of the sequences of (A)-(C). The nucleic acid is useful for
XX preparing a medicament for preventing, treating or ameliorating a medical
XX condition e.g., cancer, liver disorders such as hepatitis or neural
XX disorders such as Alzheimer's disease. This sequence encodes a novel
XX human secreted protein
XX SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;
```

```
Query Match 42.5%; Score 1054; DB 8; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 267. CTTGGCGGTATCCCGTTCGAAAGCATCATCCAGAGTTCGACCGGTGTTAACTATAG 326
Db 1. CTTGGCGGTATCCCGTTCGAAAGCATCATCCAGAGTTCGACCGGTGTTAACTATAG 60
QY 327. ATCAACACATCATCTTGCATCTCATGGGTTCTATGAATTTTAAATGGTTGATGAAG 386
Db 61. ATCAACACATCATCTTGCATCTCATGGGTTCTATGAATTTTAAATGGTTGATGAAG 120
QY 387. AGCATGGTATCCAGTAGGAAGATAGTAGGTGTTACTGTTTACCAGGGTGTGATGAAC 446
Db 121. AGCATGGTATCCAGTAGGAAGATAGTAGGTGTTACTGTTTACCAGGGTGTGATGAAC 180
QY 447. CGCTGGCCTTATTCATTTGATTTTAAATACATTAATGAACATACTCTTCAATAAGACAGT 506
Db 181. CGCTGGCCTTATTCATTTGATTTTAAATACATTAATGAACATACTCTTCAATAAGACAGT 240
QY 507. ATGTGTGTTCCCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 566
Db 241. ATGTGTGTTCCCTTGGACCAACTTTTAGCGGCTTACATCTATATCTACTTCTCTGCTTAC 300
```

Db 601 AGTCTACATAGCATATAGCACTTCTTACATTGTGGGTTTAAATATTATATCAATGCAGATACC 660
 QY 927 TTTTGTGGGATCCAGCCATCAGCAACAGTGAACATAGCGAGCTGCAGGTGCTTTTGC 986
 Db 661 TTTTGTGGGATCCAGCCATCAGCAACAGTGAACATAGCGAGCTGCAGGTGCTTTTGC 720
 QY 987 ATTGTGCAAGCTTATGCTTTCTTGTGAGTATCTGAGAGACCGGATTAACAAAACAAGATT 1046
 Db 721 ATTGTGCAAGCTTATGCTTTCTTGTGAGTATCTGAGAGACCGGATTAACAAAACAAGATT 780
 QY 1047 CAGAGCCCTTTCTTTTGGGTGATCACTAGCTGAGGTGCTGCTTCTTGTGATCAT 1106
 Db 781 CAGAGCCCTTTCTTTTGGGTGATCACTAGCTGAGGTGCTGCTTCTTGTGATCAT 840
 QY 1107 CTATTGTACTTATACAGTTTACATTCACCATGAGTGGCAGGTTTATTCTATTGGGA 1166
 Db 841 CTATTGTACTTATACAGTTTACATTCACCATGAGTGGCAGGTTTATTCTATTGGGA 900
 QY 1167 TACTGGGTATCCAAATATACATTCCTCAATATTATTGATCATGCTGAGCATCAACCTTAC 1226
 Db 901 TACTGGGTATCCAAATATACATTCCTCAATATTATTGATCATGCTGAGCATCAACCTTAC 960
 QY 1227 GACTGGGTGCTTTCTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCT 1286
 Db 961 GACTGGGTGCTTTCTTCTTTGATCTACATATCTTGTATGATCTTCCAGAGGCT 1020
 QY 1287 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1320
 Db 1021 TTGGTTCTGCATCAAAATATCAACGATGAAGA 1054

RESULT 6

AA85055

ID AAX85055 standard; DNA; 1543 BP.

XX AC AAX85055;

XX DT 30-JUL-1999 (first entry)

XX DE Human secreted protein gene No. 123.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO924836-A1.

XX PD 20-MAY-1999:

XX PF 04-NOV-1996; 98WO-US023435.

XX PR 07-NOV-1997; 97US-0064900P.

XX PR 07-NOV-1997; 97US-0064908P.

XX PR 07-NOV-1997; 97US-0064911P.

XX PR 07-NOV-1997; 97US-0064912P.

XX PR 07-NOV-1997; 97US-0064983P.

XX PR 07-NOV-1997; 97US-0064984P.

XX PR 07-NOV-1997; 97US-0064985P.

XX PR 07-NOV-1997; 97US-0064987P.

XX PR 07-NOV-1997; 97US-0064988P.

XX PR 17-NOV-1997; 97US-0066089P.

XX PR 17-NOV-1997; 97US-0066090P.

XX PR 17-NOV-1997; 97US-0066094P.

XX PR 17-NOV-1997; 97US-0066095P.

XX PR 17-NOV-1997; 97US-0066100P.

PA (HUMA-) HUMAN GENOME SCI INC.
 PI Peng P, Carter KC, Endress CA, Rosen CA, Ruben SM, Janat F;
 PI Ni J, Wei Y, Moore PA, Soppet DR, Kyaw H, Lafleur DW, Olsen HS;
 PI Shi Y, Ebner R;
 XX
 DR WPI; 1999-337740/28.
 DR P-PSDB; AAY27689, AAY27918, AAY27919, AAY27920, AAY27921, AAY27922,
 DR AAY27923.

XX New human secreted proteins and coding sequences useful for treating
 PT disorders of the immune system and hyperproliferative disorders.

PS Claim 1: Page 347; 507pp; English.

XX This sequence represents a nucleic acid molecule which encodes a secreted
 CC human protein. The gene number is given in the descriptor line. The gene
 CC can be used to generate fusion proteins by linking to the gene to a human
 CC immunoglobulin Fc portion (e.g. AAX84924) for increasing the stability of
 CC the fused protein as compared to the human protein only. The invention
 CC relates to 125 novel genes and their fragments (nucleic acid sequences;
 CC AAX84933-X85057; amino acid sequences AAY27567-127933) which are useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. Also, pathological conditions can be diagnosed
 CC by determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 125 polynucleotides, based on
 CC which tissues they are most highly expressed in (see AAX84933 for
 CC described uses)

XX SQ Sequence 1543 BP; 418 A; 299 C; 327 G; 495 T; 0 U; 4 Other;

Query Match 42.5%; Score 1054; DB 2; Length 1543;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 CTTCCGCGTCATCGCTTCGAAAGCATCTCCAGAGTTCGACCCGCTGTTAACTATAG 326

Db 1 CTTCCGCGTCATCGCTTCGAAAGCATCTCCAGAGTTCGACCCGCTGTTAACTATAG 60

QY 327 ATCAACACATCATCTTGCATCTCATGGTCTCTATGAATTTTAAATTTGTTGATGAAG 386

Db 61 ATCAACACATCATCTTGCATCTCATGGTCTCTATGAATTTTAAATTTGTTGATGAAG 120

QY 387 AGCATGGTATCCATAGCAAGATAGTAGTGTTTACCCAGGGTTGATGATAAC 446

Db 121 AGCATGGTATCCATAGCAAGATAGTAGTGTTTACCCAGGGTTGATGATAAC 180

QY 447 CGCTGGCCTTATTCATTGGATTTTAAATACATTAACATCACTGTTCCACATAAGACGT 506

Db 181 CGCTGGCCTTATTCATTGGATTTTAAATACATTAACATCACTGTTCCACATAAGACGT 240

QY 507 ATGTGTGTTCTTGCACCAACTTTTAGCGCTTACATCTATCTACTTCTCTGCTTAC 566

Db 241 ATGTGTGTTCTTGCACCAACTTTTAGCGCTTACATCTATCTACTTCTCTGCTTAC 300

QY 567 AAGAGAACTTTGGAAACCAAGGAGCAGACTTTTAGCTGCTTGTATTTGCTATTGTTAC 626

Db 301 AAGAGAACTTTGGAAACCAAGGAGCAGACTTTTAGCTGCTTGTATTTTATTTGCTACC 360

QY 627 AGGCTACATATCTCGTTCAGTGTGATCTCTTGTAAATGAAGGATGCTATTGTTTGC 686

Db 361 AGGCTACATATCTCGTTCAGTGTGATCTCTTGTAAATGAAGGATGCTATTGTTTGC 420

QY 687 ACTTCAGTTTCATCTACTTATCGGTAAATCTGTAATACTGGGTGCTAGTTTGTGGAC 746

Db 421 ACTTCAGTTTCATCTACTTATCGGTAAATCTGTAATACTGGGTGCTAGTTTGTGGAC 480

QY 747 AATGTGCTGCTTATCTATCTATATGCTCTGCTTGGGGTGGTTTATGTTATTTAT 806

Db 481 AATGTGCTGCTTATCTATCTATATGCTCTGCTTGGGGTGGTTTATGTTATTTAT 540

QY 807 CATCAATCTTATTCACCTGTCATGTTTGTGTTGTTACTGATGACAGATACAGCAAG 866

Db 628 GCTGCTGTTTATTTGCTATTGTTACAGGCTACATATCTCGGTAGTGGATCCCTT 687
Qy 661 GATATGAGGCTGCTATTTTGGCACTTCAGTTCACATACATATTATGGGTAAATCT 720
Db 688 GATAATGAGGCTGCTATTTTGGCACTTCAGTTCACATACATATTATGGGTAAATCT 747
Qy 721 GTAAAACTGGGTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 780
Db 748 GTAAAACTGGGTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 807
Qy 781 TCTGCTGCTGGGTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 840
Db 808 TCTGCTGCTGGGTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 867
Qy 841 TTACTGATGAGGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 900
Db 868 TTACTGATGAGGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 927
Qy 901 GGTTTAAATATTAATCAATGAGGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 960
Db 928 GGTTTAAATATTAATCAATGAGGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 987
Qy 961 CATGAGGCTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1020
Db 988 CATGAGGCTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1047
Qy 1021 AGAGAGGCTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1080
Db 1048 AGAGAGGCTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1107
Qy 1081 GCAGTGTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1140
Db 1108 GCAGTGTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1167
Qy 1141 AGTGGCAGGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1200
Db 1168 AGTGGCAGGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1227
Qy 1201 GCATGAGTGTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1260
Db 1228 GCATGAGTGTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1287
Qy 1261 CTGTTATGCTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1320
Db 1288 CTGTTATGCTGCTGCTATTTTGGCAATGCTGCTGCTATTCCTATTTCTATATGTC 1347
Qy 1321 GTATTTG 1327
Db 1348 GTATTTG 1354

RESULT 4
AAH33264
ID AAH33264 standard; cDNA, 1209 BP.
XX
AC AAH33264;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:320.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
FN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX

PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR P-PSDB; AAG73833.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
PS Claim 1; Page 2436; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 866 to 882 and page 7053 of the sequence listing were missing at
XX time of publication, meaning no sequences are present for SEQ ID NO:1027
XX to 1052, 7921 and 7922
SQ Sequence 1209 BP; 312 A; 222 C; 235 G; 438 T; 0 U; 2 Other;
Query Match 42.5%; Score 1054; DB 4; Length 1209;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1054; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 267 CTTGCGGCTATCCGCTCGAAAGCATCATCCAGAGTTCGACCGGTGTTAACTATAG 326
Db 1 CTTGCGGCTATCCGCTCGAAAGCATCATCCAGAGTTCGACCGGTGTTAACTATAG 60
Qy 327 ATCAACACATCATCTTGATCTCATGGTTCATGAATTTTAAATTTGTTTCAATGAAG 386
Db 61 ATCAACACATCATCTTGATCTCATGGTTCATGAATTTTAAATTTGTTTCAATGAAG 120
Qy 387 AGCATGGTATCCACTAGGAAGATAGTAGTGGTACCTTTACCAGGGTGTGATGATAC 446
Db 121 AGCATGGTATCCACTAGGAAGATAGTAGTGGTACCTTTACCAGGGTGTGATGATAC 180
Qy 447 CGCTGGCTTATTCATTTGATTTTAAATATACATTAACATTAACATTAACATTAACAT 506
Db 181 CGCTGGCTTATTCATTTGATTTTAAATATACATTAACATTAACATTAACATTAACAT 240
Qy 507 ATGTGTGTTCTTGCACCACTTTTAGCGGCTTACATCTATATCTACTTCTGCTTAC 566
Db 241 ATGTGTGTTCTTGCACCACTTTTAGCGGCTTACATCTATATCTACTTCTGCTTAC 300
Qy 567 AAGAGAACTTTTGGAAACCAAGGAGGAGGAGCTTTTAGCTGCTTGTATTATGCTATTG 626
Db 301 AAGAGAACTTTTGGAAACCAAGGAGGAGGAGCTTTTAGCTGCTTGTATTATGCTATTG 360
Qy 627 AGGTACATATCTCGTCACTAGTGGTTCCTTTGATTAAGTAAAGGAGGAGGAGGAGG 686
Db 361 AGGTACATATCTCGTCACTAGTGGTTCCTTTGATTAAGTAAAGGAGGAGGAGGAGG 420
Qy 687 ACTTCAGTTCACATATCTATTTTGGTAAATCTGTAAAAAAGTGGTCAAGTCTTTTGGAC 746
Db 421 ACTTCAGTTCACATATCTATTTTGGTAAATCTGTAAAAAAGTGGTCAAGTCTTTTGGAC 480
Qy 747 AATGTCGCTGCTTATCTTATTTCTATATGCTCTGCTGCTGCTGCTGCTGCTATTTAT 806

QY 661 GATAAAGAGGCAATGCTATTTTGGCACTCAGTCCACATACATATTTATGGTAAATCT 720
Db 688 GATAAAGAGGCAATGCTATTTTGGCACTCAGTCCACATACATATTTATGGTAAATCT 747
QY 721 GTAAAGAACTGGGTCAGTTTTTGGCACTCAGTCCACATACATATTTATGGTAAATCT 780
Db 748 GTAAAGAACTGGGTCAGTTTTTGGCACTCAGTCCACATACATATTTATGGTAAATCT 807
QY 781 TCTGCTGGGGTGGTTATGATTTATCAATCAATCTTATTCATGATGATTTATGTTG 840
Db 808 TCTGCTGGGGTGGTTATGATTTATCAATCAATCTTATTCATGATGATTTATGTTG 867
QY 841 TTACTGATGAGAGATACACAAAGAGTCTACATAGCATATAGCACCTTCTACATGTTG 900
Db 868 TTACTGATGAGAGATACACAAAGAGTCTACATAGCATATAGCACCTTCTACATGTTG 927
QY 901 GGTATTAATATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTAA 960
Db 928 GGTATTAATATCAATGAGATACCTTTTGGGATTCAGCCCAATCAGAACAGTAA 987
QY 961 CACATGGCAGTGCAGGTGCTTTTGCATTTGCTGCAAGCTTATGCTTTTCTGCAATCTG 1020
Db 988 CACATGGCAGTGCAGGTGCTTTTGCATTTGCTGCAAGCTTATGCTTTTCTGCAATCTG 1047
QY 1021 AGAGACCGATTAACAAAAGAGTTCCAGACCCCTTTTCTTTTGGGTGATCACTAGCT 1080
Db 1048 AGAGACCGATTAACAAAAGAGTTCCAGACCCCTTTTCTTTTGGGTGATCACTAGCT 1107
QY 1081 GCAGGTGCTGTTCTTCTAGTGTCACTATTTGACCTTATACAGGTATCAATGCACATGG 1140
Db 1108 GCAGGTGCTGTTCTTCTAGTGTCACTATTTGACCTTATACAGGTATCAATGCACATGG 1167
QY 1141 AGTGGAGGTTTATTCATTTGGGATCTAGGTATGCAAAATACATTTCCAAATATT 1200
Db 1168 AGTGGAGGTTTATTCATTTGGGATCTAGGTATGCAAAATACATTTCCAAATATT 1227
QY 1201 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTCTGATCTACATATT 1260
Db 1228 GCATCAGTGTCTGAGCATCAACCTACGACTTGGGTGCTTTCTTTTCTGATCTACATATT 1287
QY 1261 CTGTATGTACCTTCCAGCAGGCTTTGGTCTGATCAAAATACATTTCCAAATATT 1320
Db 1288 CTGTATGTACCTTCCAGCAGGCTTTGGTCTGATCAAAATACATTTCCAAATATT 1347
QY 1321 GTATTTG 1327
Db 1348 GTATTTG 1354

RESULT 3

ADL30661
ID ADL30661 standard; cDNA; 1664 BP.

XX AC ADL30661;

XX DT 20-MAY-2004 (first entry)

XX DE Full length human cDNA clone SeqID 2694.

XX human; medicine; signal transduction; glycoprotein; transcription;
XX oligo-capping method; ss; gene.

XX OS Homo sapiens.

XX PN EPI396543-A2.

XX PD 10-MAR-2004.

XX PF 07-JUL-2000; 2003EP-00025638.

XX PR 08-JUL-1999; 99JP-00194486.

XX PR 11-JAN-2000; 2000JP-00118774.

PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
XX Wakanatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
DR P-PSDB; ADL30662.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
XX length human cDNAs.
XX Example 1; SEQ ID NO 2694; 1340pp; English.
XX This invention relates to a novel primers useful for synthesizing full
XX length cDNA molecules that encode human proteins. Specifically, it refers
XX to secretory or membrane proteins that are potential therapeutic agents/
XX target molecules in the field of medicine, and in particular genes
XX encoding proteins that are associated with signal transduction,
XX glycoproteins and transcription. The present invention describes a method
XX for efficiently cloning a full length human cDNA from both the 5' and 3'
XX ends using the oligo-capping method. This polynucleotide sequence is a
XX full length human cDNA clone of the invention.

XX SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

Query Match 51.4%; Score 1276; DB 12; Length 1664;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCGGAGCCCTCGGCGCGGAGAGCAGCAGAGTGCCTCAACTGTCCTCCCTGG 60
Db 28 ATGGCGGAGCCCTCGGCGCGGAGAGCAGCAGAGTGCCTCAACTGTCCTCCCTGG 87
QY 61 ATGGCGCTCATGGCCCTCGGAAACAGCGCGCAGCAGCAGCAGCAGCAGCAGTGC 120
Db 88 ATGGCGCTCATGGCCCTCGGAAACAGCGCGCAGCAGCAGCAGCAGCAGCAGTGC 147
QY 121 GGCACACAGCG 180
Db 148 GGCACACAGCG 207
QY 181 GGGCTGTGCGCAGCGCGCTGGGTGCGAGTGCCTTCTCTCTTCACTCCTCTCTCT 240
Db 208 GGGCTGTGCGCAGCGCGCTGGGTGCGAGTGCCTTCTCTCTTCACTCCTCTCTCT 267
QY 241 TGGCTTGGCGGCTTCAGCTCGGCGCTCTTCCCGCTCATCGCTTCGAAAGCATCCAC 300
Db 268 TGGCTTGGCGGCTTCAGCTCGGCGCTCTTCCCGCTCATCGCTTCGAAAGCATCCAC 327
QY 301 GAGTTCGACCCCGTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGTTCAT 360
Db 328 GAGTTCGACCCCGTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGTTCAT 387
QY 361 GAATTTTAAATGGTTTGGTGAAGAGAGCATGGTATCCACTAGGAAGAAATAGTAGTGT 420
Db 388 GAATTTTAAATGGTTTGGTGAAGAGAGCATGGTATCCACTAGGAAGAAATAGTAGTGT 447
QY 421 ACTGTTTACCCAGGTTGATGATTAACCGCTGCTTATTCATTTGATTTTAAATACATTTG 480
Db 448 ACTGTTTACCCAGGTTGATGATTAACCGCTGCTTATTCATTTGATTTTAAATACATTTG 507
QY 481 AACATAACTGTTTCATTAAGAGAGCATGGTATGGTCTTCCACCACTTTTAGGGGCTT 540
Db 508 AACATAACTGTTTCATTAAGAGAGCATGGTATGGTCTTCCACCACTTTTAGGGGCTT 567
QY 541 ACATCTATATCTATCTTCTCTCTTACAGAGAACTTTTGAACCAAGAGAGAGGACTTTTA 600
Db 568 ACATCTATATCTATCTTCTCTCTTACAGAGAACTTTTGAACCAAGAGAGGACTTTTA 627
QY 601 GCTGCTGTTTATTTGCTATTTGTACCGAGCTACATATCTCGGTGAGTAGTGCCTTT 660

Db 1801 GTATGCTCTTGGGATTTATGGCTATCAGATAGCTGGAACTGGCTAATAGAACTACGTTG 1860
 Qy 1861 GTGATAATAACACCTGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATCTCTCT 1920
 Db 1861 GTGATAATAACACCTGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATCTCTCT 1920
 Qy 1921 AATGAACAGCAGCTTAAATCAATAGGAGCTCTAGATGTAGATTTGTTGTTTATT 1980
 Db 1921 AATGAACAGCAGCTTAAATCAATAGGAGCTCTAGATGTAGATTTGTTGTTTATT 1980
 Qy 1981 TTTGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGAGGTAGG 2040
 Db 1981 TTTGGAGGGTTATTGGCTATTCTGGTATGATATCAACAAATTTCTCTGAGGTAGG 2040
 Qy 2041 ATAGCTGAGGAGCAATCCCAAGACATTCGGAAGTGTGACTATTTTACCCACAGGA 2100
 Db 2041 ATAGCTGAGGAGCAATCCCAAGACATTCGGAAGTGTGACTATTTTACCCACAGGA 2100
 Qy 2101 GAATTCGGTGTAGCAAAAGCAGGATCCCTACTTTTGAATTCCTTATGATAAATG 2160
 Db 2101 GAATTCGGTGTAGCAAAAGCAGGATCCCTACTTTTGAATTCCTTATGATAAATG 2160
 Qy 2161 TCATCTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCAAGTTTGACCGA 2220
 Db 2161 TCATCTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCCAAGTTTGACCGA 2220
 Qy 2221 ACAGTAACTCTGAGATTTGGAATTAAGGACATTAATCAACATTTTGGAAAGCCTTT 2280
 Db 2221 ACAGTAACTCTGAGATTTGGAATTAAGGACATTAATCAACATTTTGGAAAGCCTTT 2280
 Qy 2281 ACATCAGACACTGGCTTTAGGATATTAAGTAAAGGACCTGTATTAACAGGAGACA 2340
 Db 2281 ACATCAGACACTGGCTTTAGGATATTAAGTAAAGGACCTGTATTAACAGGAGACA 2340
 Qy 2341 TTAGATCACAACCTCGAGTACCAACATTTTCCCAAAACAGAAATTTTGTCAAGAAG 2400
 Db 2341 TTAGATCACAACCTCGAGTACCAACATTTTCCCAAAACAGAAATTTTGTCAAGAAG 2400
 Qy 2401 ACTACCAAGAGGAGCTGGCTACATTAATAAGTGGTTTAAAGAGGCGAAGAA 2460
 Db 2401 ACTACCAAGAGGAGCTGGCTACATTAATAAGTGGTTTAAAGAGGCGAAGAA 2460
 Qy 2461 ATATCTAAGAGACTGTTTAA 2481
 Db 2461 ATATCTAAGAGACTGTTTAA 2481

RESULT 2

AAK94164

XX AAK94164 standard; cDNA; 1664 BP.

XX AC AAK94164;

XX DT 06-NOV-2001 (first entry)

XX DS Human full-length cDNA, SEQ ID NO: 2694.

XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX OS Homo sapiens.

XX PN EPI130094-A2.

XX PD 05-SEP-2001.

XX PF 07-JUL-2000; 2000EP-00114089.

XX PR 08-JUL-1999; 99JP-0019486.

XX PR 11-JAN-2000; 2000JP-00118774.

XX PR 02-MAY-2000; 2000JP-00183765.

XX PA (HELI-) HELIX RES INST.

PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI; 2001-524255/58.
 DR P-PSDB; AAM93252.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 2694; 1380pp + Sequence Listing; English.
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesizing the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a full length human cDNA of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in CD-ROM format directly
 CC from EPO.
 XX
 SQ Sequence 1664 BP; 391 A; 396 C; 411 G; 466 T; 0 U; 0 Other;

Query Match 51.4%; Score 1276; DB 4; Length 1664;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1326; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGGCGGAGCCCTCGGCGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTGG 60
 Db 28 ATGGCGGAGCCCTCGGCGCCCGGAGAGCAAGCAAGTCGTCCTCAACTCGTCCCGTGG 87
 Qy 61 AGTGCCCTCATGCGCTTGGGAAACAGCCGCAACAGCCGCGCGCGCGCGCGCGCGCG 120
 Db 88 AGTGCCCTCATGCGCTTGGGAAACAGCCGCAACAGCCGCGCGCGCGCGCGCGCGCG 147
 Qy 121 GGCACACAGGCGGCG 180
 Db 148 GGCACACAGGCGGCG 207
 Qy 181 GGGCTGTCCAGCGCGCGCTGGGTGGCAGTGGCTTCTCTCTTCCATTCCTTCTCTGGCC 240
 Db 208 GGGCTGTCCAGCGCGCGCTGGGTGGCAGTGGCTTCTCTCTTCCATTCCTTCTCTGGCC 267
 Qy 241 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATCATCCAC 300
 Db 268 TGGCTTGGCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCATCATCCAC 327
 Qy 301 GAGTTCGACCCGCTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGTCTAT 360
 Db 328 GAGTTCGACCCGCTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGTCTAT 387
 Qy 361 GAATTTTAAATTTGTTTGAATGAAGAGCATGGTATCCACTAGGAAGATAGTAGTGGT 420
 Db 388 GAATTTTAAATTTGTTTGAATGAAGAGCATGGTATCCACTAGGAAGATAGTAGTGGT 447
 Qy 421 ACTGTTTACCAGGGTGTATGATAACCGCTGGCGCTTATTCATTTGGAATTTAAATACATTG 480
 Db 448 ACTGTTTACCAGGGTGTATGATAACCGCTGGCGCTTATTCATTTGGAATTTAAATACATTG 507
 Qy 481 AACATAACTGTTCACATAAGAGAGCATGTGTGTCTCTTCCATTCGACCAACTTTAGCGCCTT 540
 Db 508 AACATAACTGTTCACATAAGAGAGCATGTGTGTCTCTTCCATTCGACCAACTTTAGCGCCTT 567
 Qy 541 ACATCTATATCTTCTCTCTCTTCAAGAGAACTTTTGGAAACCAAGAGCAGGACTTTTA 600
 Db 568 ACATCTATATCTTCTCTCTTCAAGAGAACTTTTGGAAACCAAGAGCAGGACTTTTA 627
 Qy 601 GCTGCTGTTTTATTGCTATTGTACACAGGCTACATATCTCGGTGAGTGGTCTCTTT 660
 Db 628 GCTGCTGTTTTATTGCTATTGTACACAGGCTACATATCTCGGTGAGTGGTCTCTTT 687

CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the index from GenBank.
 XX
 SQ Sequence 2481 BP; 693 A; 503 C; 577 G; 718 T; 0 U; 0 Other;
 Query Match 100.0%; Score 2481; DB 10; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGAGCCCTCGGCGCGGAGGAGCAAGCTGCTCCCTCAACTCGTCCCGTGG 60
 Db 1 ATGGCGAGCCCTCGGCGCGGAGGAGCAAGCTGCTCCCTCAACTCGTCCCGTGG 60

QY 61 AGTGGCTCATGCGCTTGGGAAACAGCGGCAACCGGCGCGGCGCGGCGCGGCGG 120
 Db 61 AGTGGCTCATGCGCTTGGGAAACAGCGGCAACCGGCGCGGCGCGGCGCGGCGG 120

QY 121 GCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
 Db 121 GCGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180

QY 181 GGGCTGTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 Db 181 GGGCTGTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

QY 241 TGGCTTGGCGGCTTCACTATGATGAAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 300
 Db 241 TGGCTTGGCGGCTTCACTATGATGAAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 300

QY 301 GAGTTCGACCGCGGCTTCACTATGATGAAAGAGCATGCTTCCGCTTCAAGAGCATAT 360
 Db 301 GAGTTCGACCGCGGCTTCACTATGATGAAAGAGCATGCTTCCGCTTCAAGAGCATAT 360

QY 361 GAAATTTTAAATTTGGTTTGGTGAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 420
 Db 361 GAAATTTTAAATTTGGTTTGGTGAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 420

QY 421 ACTGTTTACCGAGGTTGATGATGAAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 480
 Db 421 ACTGTTTACCGAGGTTGATGATGAAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 480

QY 481 AACATACTGTTTCAATGAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 540
 Db 481 AACATACTGTTTCAATGAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 540

QY 541 ACATCTATATCTATCTTCTGATGAAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 600
 Db 541 ACATCTATATCTATCTTCTGATGAAAGAGCATGCTTCCGCTTCAAGAGCATATCCAC 600

QY 601 GCTGCTTCTTTTATGCTATGTTACAGGCTACATATCTCGGTCACTAGCTGGATCCTTT 660
 Db 601 GCTGCTTCTTTTATGCTATGTTACAGGCTACATATCTCGGTCACTAGCTGGATCCTTT 660

QY 661 GATATGAGGCACTGCTATTTTGGCACTTCACTATCTATGCTGCTGCTGCTGCTGCT 720
 Db 661 GATATGAGGCACTGCTATTTTGGCACTTCACTATCTATGCTGCTGCTGCTGCTGCT 720

QY 721 GTAAAAAATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

721 GTAAAAAATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 QY 781 TCTGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 Db 781 TCTGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840

QY 841 TTACTGATGCGAGATACAGCAAGAGCTGCTACATAGCATATAGCACTTCTACATGCTG 900
 Db 841 TTACTGATGCGAGATACAGCAAGAGCTGCTACATAGCATATAGCACTTCTACATGCTG 900

QY 901 GGTTTAATATATATCAATCCAGATACCTTTTGTGGATTCAGCCAAATCAGAAAGTGA 960
 Db 901 GGTTTAATATATATCAATCCAGATACCTTTTGTGGATTCAGCCAAATCAGAAAGTGA 960

QY 961 CACATGGGAGCTGCGAGGCTGCTTGGCATGCTGCAAGCTTATGCTTCTTGGAGTATCTG 1020
 Db 961 CACATGGGAGCTGCGAGGCTGCTTGGCATGCTGCAAGCTTATGCTTCTTGGAGTATCTG 1020

QY 1021 AGAGCCGATTAACAAAACAGAGCTTCCAGACCTTTTCTTTTGGGTGATCAGTAGCT 1080
 Db 1021 AGAGCCGATTAACAAAACAGAGCTTCCAGACCTTTTCTTTTGGGTGATCAGTAGCT 1080

QY 1081 GCAGGTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140
 Db 1081 GCAGGTGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1140

QY 1141 AGTGGCAGGCTTTATTCATTTGGGATATCGGGTATCGAAATATACATTCCTCAATAT 1200
 Db 1141 AGTGGCAGGCTTTATTCATTTGGGATATCGGGTATCGAAATATACATTCCTCAATAT 1200

QY 1201 GCATCAGTCTGTCAGCATCAACCTACAGCTTGGGTGCTTCTTCTTCTTCTTCTTCTTCT 1260
 Db 1201 GCATCAGTCTGTCAGCATCAACCTACAGCTTGGGTGCTTCTTCTTCTTCTTCTTCTTCT 1260

QY 1261 CTTGTATGTACCTTCCAGCGGCTTTGGTCTGCAATCAAAATATACACGATGAAGA 1320
 Db 1261 CTTGTATGTACCTTCCAGCGGCTTTGGTCTGCAATCAAAATATACACGATGAAGA 1320

QY 1321 GTATTTGTTGCTCTATATGCAATCAGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
 Db 1321 GTATTTGTTGCTCTATATGCAATCAGTCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380

QY 1381 ATGTTGACTTTGACTTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 Db 1381 ATGTTGACTTTGACTTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

QY 1441 GAGCCTATTTGGGGGATGACATGAAAGAGGAAATCCACCTGTGGAGGACAGCAGTGAT 1500
 Db 1441 GAGCCTATTTGGGGGATGACATGAAAGAGGAAATCCACCTGTGGAGGACAGCAGTGAT 1500

QY 1501 GAGGATGACAAAGAAACAGGAAATTTGTATGATGAGGAGGAAATTTGTATGAGGAGGAA 1560
 Db 1501 GAGGATGACAAAGAAACAGGAAATTTGTATGATGAGGAGGAAATTTGTATGAGGAGGAA 1560

QY 1561 GCACTGAAACAGGAAACAACTGAAGAGGATTTAGGCGCTTAAATATAAAGCAATGTCACC 1620
 Db 1561 GCACTGAAACAGGAAACAACTGAAGAGGATTTAGGCGCTTAAATATAAAGCAATGTCACC 1620

QY 1621 ATGTTGATGCTGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
 Db 1621 ATGTTGATGCTGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680

QY 1681 GCTACTCTAGTCCAGTGTAGTCTGCTCCTCATCAATCATGATGCAACAGAAATATC 1740
 Db 1681 GCTACTCTAGTCCAGTGTAGTCTGCTCCTCATCAATCATGATGCAACAGAAATATC 1740

QY 1741 TTAGATGATTTTATGAGAGAGCTTCTTTTGGCTTAAAGGAAATACAGATGAACATGACGA 1800
 Db 1741 TTAGATGATTTTATGAGAGAGCTTCTTTTGGCTTAAAGGAAATACAGATGAACATGACGA 1800

QY 1801 GTAAATGCTGTTGGGATTTAGGCTATCAGATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 15, 2004, 00:53:33 ; Search time 1667 Seconds
(without alignments)
7812.721 Million cell updates/sec

Title: US-10-028-384-1
Perfect score: 2481
Sequence: 1 atggcgagccctcgcccc.....tatctaagaagactgttaa 2481

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_23Sep04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 2481 | 100.0 | 2481 | 10 | ADD94783 |
| 2 | 1276 | 51.4 | 1664 | 4 | AAK94164 Human ful |
| 3 | 1276 | 51.4 | 1664 | 12 | ADL30661 Full leng |
| 4 | 1054 | 42.5 | 1209 | 4 | AAH33284 Human col |
| 5 | 1054 | 42.5 | 1209 | 6 | ABL89850 Human pol |
| 6 | 1054 | 42.5 | 1543 | 2 | AAK85055 Human sec |
| 7 | 1054 | 42.5 | 1543 | 8 | ACD18981 Novel hum |
| 8 | 1054 | 42.5 | 1543 | 12 | ADG78372 Human sec |
| 9 | 1054 | 42.5 | 1543 | 12 | ADN60663 Human sec |
| 10 | 972 | 39.2 | 2546 | 2 | AAV44866 Clone CTS |
| 11 | 972 | 39.2 | 2546 | 5 | AAK98463 Human CDN |
| 12 | 954 | 38.5 | 2510 | 5 | AAK93772 Human CDN |
| 13 | 891 | 35.9 | 2537 | 4 | AAK08289 Human sec |
| 14 | 861 | 34.7 | 2284 | 4 | AAH18021 Human CDN |
| 15 | 840 | 33.9 | 2547 | 4 | AAK08315 Human sec |
| 16 | 840 | 33.9 | 2660 | 6 | ABQ54750 Human ova |
| 17 | 653 | 26.3 | 1114 | 4 | AAH99794 Human pro |
| 18 | 521 | 21.0 | 764 | 5 | AAK93968 Primer sp |
| 19 | 507 | 20.4 | 507 | 3 | AAK01967 Human sec |
| 20 | 506 | 20.4 | 787 | 4 | AAH07526 Human CDN |
| 21 | 500 | 20.2 | 500 | 6 | ABV88332 Human col |

| | | | | | |
|----|-----|------|------|----|--------------------|
| 22 | 430 | 17.3 | 433 | 6 | ABL37128 Human col |
| 23 | 402 | 16.2 | 558 | 5 | AAK93612 Umbilical |
| 24 | 359 | 14.5 | 387 | 4 | ABA08397 Human sec |
| 25 | 351 | 14.1 | 616 | 3 | AAZ80248 Human col |
| 26 | 328 | 13.2 | 485 | 9 | ACH24071 Human adu |
| 27 | 322 | 13.0 | 349 | 2 | AAV89737 EST clone |
| 28 | 285 | 11.5 | 313 | 8 | ABZ20328 Group III |
| 29 | 250 | 10.1 | 250 | 3 | AAK01440 Human sec |
| 30 | 236 | 9.5 | 245 | 3 | AAK17015 Human sec |
| 31 | 236 | 9.5 | 247 | 2 | AAK40252 Human sec |
| 32 | 231 | 9.3 | 281 | 3 | AAK45221 Human sec |
| 33 | 218 | 8.8 | 503 | 4 | AAK91644 Human CDN |
| 34 | 218 | 8.8 | 503 | 4 | AAK93577 Human CDN |
| 35 | 218 | 8.8 | 503 | 12 | ADL30004 |
| 36 | 218 | 8.8 | 503 | 12 | ADL28071 |
| 37 | 188 | 7.6 | 388 | 5 | AAK66328 |
| 38 | 184 | 7.4 | 737 | 4 | AAI96771 |
| 39 | 179 | 7.2 | 474 | 5 | AAK93633 |
| 40 | 156 | 6.3 | 393 | 4 | ABA08547 Human sec |
| 41 | 124 | 5.0 | 355 | 6 | ABL37443 Human col |
| 42 | 114 | 4.6 | 567 | 4 | AAK92463 Human CDN |
| 43 | 114 | 4.6 | 567 | 12 | ADL28890 |
| 44 | 105 | 4.2 | 483 | 9 | ACH36684 Human end |
| 45 | 65 | 2.6 | 2710 | 10 | ADD94785 Mouse SIM |

ALIGNMENTS

RESULT 1
ADD94783
ID ADD94783 standard; cDNA; 2481 BP.

AC ADD94783;
XX
XX
DT 29-JAN-2004 (first entry)
XX
DE Human SIMP cDNA sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
KW gene; ss.
XX Homo sapiens.
XX
XX WO2003054008-A2.
XX
XX 03-JUL-2003.
XX
XX 18-DEC-2002; 2002WO-CA001967.
XX
XX 20-DEC-2001; 2001US-00028384.
XX
XX (COMP-) COMPATIGENE INC.
XX
XX Perreault C, McBride K;
XX
XX WPI, 2003-559122/52.
XX
XX F-PSDB; ADD94784.

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Claim 6; SEQ ID NO 1; 66pp; English.

This invention relates to a novel isolated or purified human protein,

| | | | |
|----|------|---|------|
| Db | 1552 | ATTGACGATTATCGTGAGGCTTACTACTGGCTTCGTGAGATAATCTCTCAGAACGCCAAA | 1611 |
| Qy | 601 | ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu | 620 |
| Db | 1612 | ATCATGTGATGGTGGGATTATGGGTATCAATTCGGTGCATGGCGGACCGCCAAACCTTG | 1671 |
| Qy | 621 | ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer | 640 |
| Db | 1672 | GTTGACCAACACACCTGGACCAACACCCATATGTACGGTGGTGAAGGCGATGAGTCGA | 1731 |
| Qy | 641 | AsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAlaAspTyrValLeuValIle | 660 |
| Db | 1732 | CGCGAGGAAGTCAGTACCCATCTCCGCCAGCATGATGCGATTACGTGCTGGTGGTG | 1791 |
| Qy | 661 | PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTpmMetValArg | 680 |
| Db | 1792 | TTCGGTGGTCTGTAGGTATTCTGCGGATGACATTAAACAATTCATTATGATGGTCCGT | 1851 |
| Qy | 681 | IleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGly | 700 |
| Db | 1852 | ATCGCCGAAGGTATCTGGCCGATGAGGTAAAGAGCGGGACTCTTTACTGCACGCGGT | 1911 |
| Qy | 701 | GluPheArgValAlaPylsAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLysMet | 720 |
| Db | 1912 | GAATTCGTGTCCAGCATGAGCGACCCCACTATGCGCAACAGCTTGTGGTATAAATG | 1971 |
| Qy | 721 | SerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAspArg | 740 |
| Db | 1972 | TCATTATCAATTCAACTCTCTC-----TTCCCGTCGGGCCAAGTGTGCACGCG | 2022 |
| Qy | 741 | ThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPhe | 760 |
| Db | 2023 | GTCCGTGGGTCAAACATTCCACAGAGGCCCTCAGCTCTCTACACTCGAAGAGACTTTC | 2082 |
| Qy | 761 | ThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGluThr | 780 |
| Db | 2083 | ACGAGCGAGAAGCTGGATCATTCGTATCTACAAAGGTCAAGGATCTTGACAACTTGGCCGA | 2142 |
| Qy | 781 | LeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLys | 800 |
| Db | 2143 | ---GACCACCAACGAGCTGTGCC-----TTGCACAAGGTTCTCAAGAAAAG | 2187 |
| Qy | 801 | ThrThrLysArgLysArgGly | 807 |
| Db | 2188 | CGGAGTACAAAGAGGAAGGA | 2208 |

```

RESULT 15
US-10-032-585-6323
/ Sequence 6323, Application US/10032585
/ Publication No. US20030180953A1
/ GENERAL INFORMATION:
/ APPLICANT: Terry, Roemer D.
/ APPLICANT: Bo, Jiang
/ APPLICANT: Charles, Boone
/ APPLICANT: Howard, Bussey
/ TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
/ FILE REFERENCE: 10182-005-999
/ CURRENT APPLICATION NUMBER: US/10/032,585
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 8000
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 6323
/ LENGTH: 2256
/ TYPE: DNA
/ ORGANISM: Candida albicans
US-10-032-585-6323

```

| | |
|------------------------|-----------|
| Alignment Scores: | |
| Pred. NO.: | 3.35e-214 |
| Score: | 2164.00 |
| Percent Similarity: | 71.39% |
| Best Local Similarity: | 56.15% |
| Percent Similarity: | 51.11% |
| Mismatches: | 176 |
| Conservative: | 114 |
| Matches: | 420 |
| Length: | 2256 |


```
Db      2116 ATTGTGAGAAATTACAAAGTTAAAGATTTCGATAATGTTGGTAGAGATTTCACATCAA--- 2172
Qy      786 ArgValThrAsnIlePheProLysGlnLysTyrLeuSerLysLysThrThrLysArgLys 805
Db      2173 -----GCTACTGCTTTTGAAGATCA-----TCATCCGGCAGCTTCCAAAGAAAC 2217
Qy      806 ArgGlyTyrIleLysAsnLysLeu 813
Db      2218 AGATCCATAAAGAGACCTAAATTG 2241
```

Search completed: December 15, 2004, 21:14:11
Job time : 1185.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 6394.33 seconds
(without alignments)
4707.175 Million cell updates/sec

Title: US-10-028-384-2

Perfect score: 4401

Sequence: 1 MAEPSAPSRKHSKSLNSPPW.....GYINKLVFKKIKKTV 826

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q/cg2_1/USPIO_spool/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -DOOPCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.13960 @runat_14122004_131519_8030 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPEXT=10 -XGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsal.*
9: gb_gsal2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|------------|--------------------|
| 1 | 4263.5 | 96.9 | 2709 | 3 AK018758 | AK018758 Mus muscu |
| 2 | 4216.5 | 95.8 | 2669 | 3 AK012153 | AK012153 Mus muscu |
| 3 | 3918 | 89.0 | 4513 | 3 BC044321 | BC044321 Xenopus 1 |
| 4 | 2763 | 62.8 | 2334 | 3 AY310157 | AY310157 Rattus no |
| 5 | 2259.5 | 51.3 | 3802 | 3 AK030363 | AK030363 Mus muscu |
| 6 | 2258.5 | 51.3 | 2656 | 3 BC028897 | BC028897 Mus muscu |
| 7 | 2255.5 | 51.2 | 2730 | 3 AK081547 | AK081547 Mus muscu |
| 8 | 2238 | 50.9 | 3643 | 3 AK077877 | AK077877 Mus muscu |
| 9 | 2134 | 48.5 | 2118 | 9 AY418284 | AY418284 Homo sapi |

| | | | | | |
|----|--------|------|------|------------|--------------------|
| 10 | 2112 | 48.0 | 2727 | 3 AK087470 | AK087470 Mus muscu |
| 11 | 2103.5 | 47.8 | 2103 | 9 AY418286 | AY418286 Mus muscu |
| 12 | 2065.5 | 46.9 | 2419 | 3 CNS0ADY2 | AY418285 Arabidops |
| 13 | 1691 | 38.4 | 2079 | 9 AY418285 | AY418285 Pan trogl |
| 14 | 1563 | 35.5 | 950 | 5 BU515954 | BU515954 AGENCOURT |
| 15 | 1513 | 34.4 | 2212 | 3 CR610915 | CR610915 full-len |
| 16 | 1465 | 33.3 | 852 | 7 CN458732 | UI-M-H30 |
| 17 | 1465 | 33.3 | 879 | 5 BQ947308 | AGENCOURT |
| 18 | 1391 | 31.6 | 823 | 7 CF736482 | UI-M-HD0 |
| 19 | 1383 | 31.4 | 863 | 5 BU911227 | AGENCOURT |
| 20 | 1361 | 30.9 | 867 | 5 E1851585 | E1851585 603378796 |
| 21 | 1351 | 30.7 | 798 | 7 CN527592 | UI-M-HQ0 |
| 22 | 1337 | 30.4 | 1088 | 5 BX363211 | BX363211 BX363211 |
| 23 | 1333 | 30.3 | 766 | 7 CK633751 | CK633751 UI-M-HQ0 |
| 24 | 1329.5 | 30.2 | 814 | 7 CN531243 | UI-M-HN0 |
| 25 | 1324.5 | 30.1 | 896 | 6 CA980464 | CA980464 AGENCOURT |
| 26 | 1318 | 29.9 | 957 | 5 BU554374 | BU554374 AGENCOURT |
| 27 | 1313 | 29.8 | 825 | 6 CA318063 | CA318063 UI-M-FM0 |
| 28 | 1302.5 | 29.6 | 807 | 5 BU132963 | BU132963 60311925 |
| 29 | 1291 | 29.3 | 765 | 6 CA382416 | CA382416 662214 NC |
| 30 | 1289 | 29.3 | 735 | 4 BG698273 | BG698273 602660188 |
| 31 | 1289 | 29.3 | 754 | 7 CK636437 | CK636437 UI-M-HQ0 |
| 32 | 1279 | 29.1 | 724 | 7 CF737921 | CF737921 UI-M-HD0 |
| 33 | 1276 | 29.0 | 877 | 6 CB208902 | AGENCOURT |
| 34 | 1270.5 | 28.9 | 775 | 7 CF730893 | UI-M-GZ0 |
| 35 | 1265.5 | 28.8 | 821 | 7 CF578043 | CF578043 AGENCOURT |
| 36 | 1261 | 28.7 | 738 | 7 CNE27774 | UI-M-HQ0 |
| 37 | 1256.5 | 28.6 | 861 | 7 CF995035 | CF995035 AGENCOURT |
| 38 | 1255 | 28.5 | 960 | 6 CD793779 | CD793779 EST665140 |
| 39 | 1252 | 28.4 | 759 | 4 B1662821 | B1662821 603286252 |
| 40 | 1249 | 28.4 | 724 | 5 BP701425 | BP701425 BP701425 |
| 41 | 1249 | 28.4 | 742 | 5 BQ443871 | UI-M-EW0 |
| 42 | 1243.5 | 28.3 | 798 | 7 CK016163 | CK016163 AGENCOURT |
| 43 | 1243 | 28.2 | 721 | 7 CNE29521 | UI-M-HQ0 |
| 44 | 1242.5 | 28.2 | 876 | 6 CD794446 | EST65807 |
| 45 | 1240.5 | 28.2 | 779 | 5 BM963786 | UI-M-EQ0 |

ALIGNMENTS

AK018758 2709 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:130006C19 product:hypothetical oligosaccharyl transferase (OTase) STT3 subunit containing protein, full insert sequence.

ACCESSION AK018758
VERSION AK018758.2 GI:26384577
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

REFERENCE 2 10349636

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4
FANTOM Consortium.
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 2709)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan. (E-mail: genome-res@gsr.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

On Dec 10, 2002 this sequence version replaced gi:12858635.
Please visit our web site (<http://genome.gsc.riken.jp/>) for further
details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGAGCATCAGAGCATCAATTAATTAACCCCGCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

Location/Qualifiers
1..2709
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:1300006C19"
/db_xref="taxon:10090"
/clone="1300006C19"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
76..2547

CDS
note="unnamed protein product; hypothetical
Oligosaccharyl transferase (OTase) S7r3 subunit containing
protein (InterPro|IPR003674, evidence: InterPro)
putative"
/codon_start=1

JOURNAL

US-10-028-384-2 (1-826) x AK018758 (1-2709)

Qy 1 MetAlaGluProSerAlaProGluSerIshHisIshSerSerLeuAenSerSerProTrp 20

Db 76 ATGGCGGAGCCCTCGGCCCGGAGCAAGCAAGTCCTCCCTCACTGCTCCCGTGG 135

Qy 21 SerGlyLeuMetAlaLeuGlyValSerArgHisGlyHisGlyProGlyAlaGlnCys 40

Db 136 AGCGGCTCATGGCTCTGGGGAACAGCGCCAGGCGCATGGCCCGGAGCCAGAGC 195

Qy 41 AlaHisLysAlaAlaGlyGlyAlaAlaProProLysProAla---ProAlaGlyLeuSer 59

Db 196 GCGTCCAGCGCGCGCGCGC-----CCGAGCGCGCGCGCGCGCGCGCGTGTCC 243

Qy 60 GlyGlyLeuSerGlnProAlaGlyTrpGlnSerLeuLeuSerPheThrLeuPheLeu 79

Db 244 GGGGCTTGTCCAGCG 303

Qy 80 AlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValleArgPheGluSerIlelle 99

Db 304 GCGTGGTGGCGGCTTCAGCTCGCGCTCTTCGCGCGCGCGCGCGCGCGCGCGCG 363

Qy 100 HisGluPheAspProTrpPheAsnTyrArgSerThrHisIshLeuAlaSerHisGlyPhe 119

Db 364 CACGAGTTCGACCGCGTGTAACTATAGATCAACATCATCTTGCATCTTCATGATTC 423

Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139

Db 424 TATGAGTTCTAAATGTGTGATGAAAGAGCATGGTACCCACTGGGAAAGATAGTGGGT 483

Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThr 159

Db 484 GGCACCGTTACCCAGCGGTTCATGATCAACAGCTGGCCTTATTTCATTTGATATACA 543

Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179

Db 544 TTGAACATCAACAGTTTCATAGAGATGTGTGTGTATTCCTTGCACCACTTTAGCGGC 603

Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeu 199

Db 604 CTTACATCATATCTACGTTCTCTCTAAGTAGAGAACTGTGGAAACCAAGGAGGAGACTT 663

/protein_id="BAB1390.1"

/db_xref="GI:12858636"

/translation="MAEPAPESKHSSLNSSPWSGLMALGNSRHGHGHGPGTOSASRA
AAPKPPAGLSGGLSOPAGWQSLSTLFLAWLAGFSRLFAVIRFESIHEPDPW
FYHRTVHLASHGFVEFLNFDERAWYPLRGVTVGLMIMTAGLHILNLTNIT
NHRDVCVLAFTFSLTSTISFTLLRELNOGAGLLAACFIAIVPGYISRSVAGSFD
NEGIAIPALQTYVYLMVKSQSVFVTCCLSYFVMSAGWGVFIINLILPHVFD
LLLMORYSKRYVYIAYSTLYIVGLILSNQIPVPGFQPIRSEHMAAGVAFALQAFAL
QIQRDLTKQBFQTLFLGVSAAAGAVLSVITITGYITAPWSGFRYSFISLWDITGYAKI
HIPVILASVSEHQPTTWVSFFDLVLTCPAGLFCIKNINDESRVFAVIAISPGNLY
AGVIMLMTLTTPVCMLSAIAFNSVHYGLDDMKENPVEDSDDEKRNPNGLY
DKAGVRKHVTEQKEPBEGLGNFKISVTMLMLMLMFAVHCTVWTNAYSPPSYVL
ASYNHGTRNLDLDFREAYFWLQNTDHEARVMSWMDYQIAGMANRTTLVNNNTWN
NSHIALVGMKSNSETRAYKIMRSLDVIDVYLVIFGVIGYGGDDINKFLMVMVIRABGE
HKPIDREGDYFTQGEPRVDKAGSPITLNLCLMYKMSYRFGENCLDFRTPGPDTRN
AELGNKDIKFHLSEATSEHLVRIYKVAFDNRETLGHKPRVTNIVPKQVLSKKT
TKRKGIVKGLVFKKTKSKIV"

2693..2698

/note="putative"

2709

/note="putative"

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2709

Score: 4263.50 Matches: 805

Percent Similarity: 97.70% Conservativeness: 3

Best Local Similarity: 97.34% Mismatches: 14

Query Match: 96.88% Indels: 5

DB: 3 Gaps: 2

200 LeuAlaAaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
 Db CTAGCTGCTGCTTCATTGCTATCGTACAGGGTACATATCTCGTCAAGTGGCGGATCC 723
 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLys 239
 Db TTTGATATGAAGGCAATGGCAATTTTGGCGTTCAGTTCACTTACTTATGGGTAAG 783
 240 SerValLysThrGlySerValPheTrpThrMetCysCysCysLeuSerTyrPheTyrMet 259
 Db TCTGTGAAGACCGGCTGCTGCTTCTGGACAAATGCTGCTGCTTGTCAATTTCTCATG 843
 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
 Db GTCTCTGCGTGGGAGGTTATGTTTCAATCAACCTCACTCCCTCCAGTGTGTTG 903
 280 LeuLeuLeuMetGlnArgTyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIle 299
 Db TTGCTGCTGATGCGAGGTACAGCAAGAGAGTCTACATAGCATATAGCACTTTGTACAT 963
 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
 Db GTGGGTTTAAATATATCATGCAATACCTTTTGTGGGATTCAGCAATCAAGAACAGC 1023
 320 GluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
 Db GAGCACATGGCAGCTGCAGGTGCTTCTTGGCGTGTGCAAGCTTACGCTTTTTCAGTAT 1083
 340 LeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPheLeuGlyValSerLeu 359
 Db CTGAGAGACCGGTGACAAAACAGAGTTCCAGACCCCTTTCTTTTGGGTGCTCACA 1143
 360 AlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPro 379
 Db GCTGAGCGCTGCTGCTTCTTAGTGTCTATCTATCTGACATACAGGTTATATGACCA 1203
 380 TrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle 399
 Db TGGAGTGCAGGTTTATCTACTATGGATCTGGTATCGGATATGCAAAATATACATTCAT 1263
 400 IleAlaSerValSerGluHisGlnProThrTrpValSerPhePheAspLeuHis 419
 Db ATTGTCATCAGTGTCTGAACATCAGCCTACGACATGGGTGCTTCTTCTTTGATCTACAT 1323
 420 IleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGlu 439
 Db ATTCTTGTATGTACTTCCAGCAGGCTATGTTCTGATCAATAATATCAACGATGA 1383
 440 ArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 459
 Db AGAGTATTGTGCTCTGTATGCGATCAGTGTCTGTACTTTGCGGAGTGTATGGTGGCG 1443
 460 LeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVal 479
 Db CTGATGCTGACTCTGACCCCGGTGCTGTGATGCTGTGCGCCATCGCTTCTCCATGTT 1503
 480 PheGluHisTyrLeuGlyAspMetLysArgGluAsnProValGluAspSerSer 499
 Db TTTGAGCCTATTGGGGGATGACATGAAGAGGAAACCCACCTGTGGAGGACAGCAGT 1563
 500 AspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLys 519
 Db GATGAGGATGACAAAGAAACCCAGGAAACCTTGTATGACAGGAGGATTAAGTGAAGAG 1623
 520 HisAlaThrGluGlnGluTyrGluGluGlyLeuGlyProAsnIleLysSerIleVal 539
 Db CATGTGACAGAGCAAGAAACCTGAAGAGGCTTGGGGCCCAACATCAAAAGCATGTG 1683
 540 ThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThrTrpValThrSer 559
 Db ACCATGCTGATGCTCATGCTCTCTGATGATGTTTCCGGGTCCACTGCAGGTGGTCAACAGC 1743
 560 AsnAlaTyrSerSerProSerValIleLeuAlaSerTyrAsnHisAspGlyThrArgAsn 579

1744 AACGCCCTACTCCAGTCCAGTGTGGTCTCTCTCAATCATGATGATGATGATGATGAT 1803
 580 IleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAla 599
 1804 ATATTAGATGATTTAGAGAGGCTACTTTTGGCTGAGACAAACACACGATGACACAGC 1863
 600 ArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThr 619
 1864 CGGTCATGTCGTGGTGGGACTACGCTATCAGATTGCTGGCATGGCCAAACAGGACCACT 1923
 620 LeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSer 639
 1924 CTGCTGATTAACACACCTGGACACAGCCACATCGCACTGGTGGGAAAGCTATGCT 1983
 640 SerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVal 659
 1984 TCCAATGAACCGCGCCCTATAAAATCATGAGTCCCTTGTATGATGATGATGATGATG 2043
 660 IlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVal 679
 2044 ATTTTCCGAGGAGTATGGCTATTCGGGGAGCATATCAACAGTTCCTCTGGATGGTC 2103
 680 ArgIleAlaGluGlyGlnHisProLysAspIleArgGluSerAspTyrPheThrProGln 699
 2104 AGGATAGCTGAAGGGGAGCATCCCAAGACATCCCGGAAAGGTGACTATTTCCACCCAGCAG 2163
 700 GlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLys 719
 2164 GGAGAGTTCGAGTAGACAAAGCTGGTCTCTCTACTCTGTTAACTGCTTATGATATAA 2223
 720 MetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAsp 739
 2224 ATGTCATACTACAGATTGGAGAAATGCACTAGATATTTCCGACTCCCGACGCTTTGAC 2283
 740 ArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluAla 759
 2284 CGAACAGTAAATGCTGAGATTGGAAATTAAGACATTAATTAATTAATTAATTAATTA 2343
 760 PheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGlu 779
 2344 TTTACATCAGAGCAGCTGGCTTGTGAGATATATAAAGTGAAGACCTGACAAACAGGAG 2403
 780 ThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLys 799
 2404 ACATAGGTCAACAACTCGAGTCACCAACATCGTCCCAACACAGAGTATTTGTCAAG 2463
 800 LysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLys 819
 2464 AAGACTACTAAAGGAGCGTGGCTAGCTAGTAAATAAAGTAAAGTAAAGTAAAGTAAAG 2523
 820 LysIleSerLysLysThrVal 826
 2524 AAGACCTTAAGAGAGCTGTT 2544

RESULT 2

AK012153

LOCUS

DEFINITION

MUS MUSCULUS 10 DAYS EMBRYO WHOLE BODY CDNA, RIKEN FULL-LENGTH

OLIGONUCLEOTIDE LIBRARY, CLONE:2610524N02 PRODUCT:HYPOTHELICAL

FULL INSERT SEQUENCE.

ACCESSION

AK012153.1

VERSION

HTC; CAP TRAPPER.

KEYWORDS

MUS MUSCULUS (HOUSE MOUSE)

SOURCE

MUS MUSCULUS

ORGANISM

EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; EUTELEOSTOMI;

MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.

REFERENCE

1

CARNINCI P. AND HAYASHIZAKI Y.

HIGH-EFFICIENCY FULL-LENGTH CDNA CLONING

METH. ENZYMOL. 303, 19-44 (1999)

| | | | |
|----|------|---|------|
| Db | 683 | TTTGATAATGAAGCATTGCCATTTTGGCGCTTCAGTTCACTTACTACTTATGGGTAAAG | 742 |
| Qy | 240 | SerValIysThrGlySerValPheTrpThrMetCysCysLeuSerTyrPheTyrMet | 259 |
| Db | 743 | TCTGTGAAGACCGGGTCTGTCTCTGGACAATGTGCTGTGTTCATATTCTTCACATG | 802 |
| Qy | 260 | ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal | 279 |
| Db | 803 | GTCTCTGGCTGGGGAGGTATGTGTTTCATCATCAACTCATCCCTCTCATGTGTGTGG | 862 |
| Qy | 280 | LeuLeuLeuMetGlnArgTyrSerIysArgValTyrIleAlaTyrSerThrPheTyrIle | 299 |
| Db | 863 | TTGTCTCTCATGCAGAGGTACAGACAAGAGAGTCTCATATAGCATATAGCACTTCTACATT | 922 |
| Qy | 300 | ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer | 319 |
| Db | 923 | GTGGGTTTAATATTATCCATACGATACCTTTTGTGGGATTTTCAGCCAATCAGCACAGC | 982 |
| Qy | 320 | GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr | 339 |
| Db | 983 | GAGCACATGCCAGCTGCAGGTGTCTTTGGCTGCTGCAAGGGTTACGGCTTTTTCGAGTA | 1042 |
| Qy | 339 | rIleuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLe | 359 |
| Db | 1043 | TCTGAGAGHACCCGGTTGACAAAACAGAGATTCACAGACCTTTTCTTTTGGGTGTCCTACT | 1102 |
| Qy | 359 | uAlaAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaPr | 379 |
| Db | 1103 | AGCTGCAGCGCGTGTGTCTTCCTTAGTGTCACTATCTGCATACACAGGTTATATTGCACC | 1162 |
| Qy | 379 | oTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIle | 399 |
| Db | 1163 | ATGGAGTGGCAGGTTTATTCACTATGGGATCTGGGTATGCAAAATACACATTCACAT | 1222 |
| Qy | 399 | eIleAlaSerValSerGluHisGlnProThrThrTrpValSerPhePheAspLeuHi | 419 |
| Db | 1223 | TATTGCATCAGTGTCTGAACATCAGCCTAGCACATGGGTGTCTTTCTCTTTGATCTACA | 1282 |
| Qy | 419 | sIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGly | 439 |
| Db | 1283 | TATTCTGTATGTACTCTCCACGACGGCCTATGGTCTGCNTCAAAATATCAACGATGA | 1342 |
| Qy | 439 | uArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValAr | 459 |
| Db | 1343 | AAGAGTATTTTGTTCGTCTGTATGGGATCAGTGTCTACTTTTGGCGGAGTGTGGTGGC | 1402 |
| Qy | 459 | gIleuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAlaPheSerAsnVa | 479 |
| Db | 1403 | GCTGATGCTGACTCTGACCCCGGTCTGTGCATGCTGTGGGCCATCGCCTTCTCCATGT | 1462 |
| Qy | 479 | lPheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProProValGluAspSerSe | 499 |
| Db | 1463 | TTTGTGACACTATTTGGGGGTGACATGAAGAAGGAAAAACCCACTGTGTGGAGGACAGC | 1522 |
| Qy | 499 | rAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgGly | 519 |
| Db | 1523 | TGATGAGGATGACAAAGAAGAACCCAGGAACCTTGTATGACAGCCGCGGTAAAGTGAGAA | 1582 |
| Qy | 519 | sHisAlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAsnIleLysSerIleVa | 539 |
| Db | 1583 | GCAATGTGACAGACGAAGAAACCTGGAAGAGGGCTTGGGCCCCAACATCAAAAGCATTTGT | 1642 |
| Qy | 539 | lThrMetLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSe | 559 |
| Db | 1643 | GACCATGCTGATGCTCATGCTCCTGATGATGTTTCGGGGTCCACTGCACGTGGGTACACAG | 1702 |
| Qy | 559 | rAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAs | 579 |
| Db | 1703 | CAACGGCTACTCCAGTCCAAGTGGTCCCTTGGCTCCTCAATCATATGATGGTACAGGAA | 1762 |
| Qy | 579 | nIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThrAspGluHisAl | 599 |

| | | | |
|------------|--|--|-----------------|
| Db | 1763 | TATATTAGATGATTTTAGAGAGCGTACTTTTGGCTGAGACAAACACAGGATGACACGC | 1822 |
| Qy | 599 | aArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrTh | 619 |
| Db | 1823 | CSGGGTCTATGTCGTGGGACCTACGGCTATCATAGATTGCTGGCATGGCCACACAGCACCAC | 1882 |
| Qy | 619 | rLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSe | 639 |
| Db | 1883 | TCTGGTGGATAAACAACACCTGGAAACAACAGCACATCGCACTGGTCGGAAGAAGCTATGTC | 1942 |
| Qy | 639 | rSerAsnGluThrAlaIalaTyrLysIleMetArgThrLeuAspValAspTyrValLeuVa | 659 |
| Db | 1943 | TTCCAAATGAACGGCGCGCTATAAAATCATGAGGTCCTTGAATGTCGATTATGTGTGGT | 2002 |
| Qy | 659 | lIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetVa | 679 |
| Db | 2003 | TATTTTCGGAGGAGTGATTGGCTATTCCGGGGACGATATCAACAAGTAGTTCTCTCGATGGT | 2062 |
| Qy | 679 | lArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrProGl | 699 |
| Db | 2063 | CAGGATAGCTGACGGGGGNGCATCCCAAGACATCCGGGAAGTGACTATTTCACCCAGCA | 2122 |
| Qy | 699 | nglyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyrLy | 719 |
| Db | 2123 | GGGAGAGTCCGAGTAGACAAAGCTGGGTCTCTACTCTGTAAACTGCCATTATGTATAA | 2182 |
| Qy | 719 | sMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPheAs | 739 |
| Db | 2183 | AATGTCATACATACAGATTTTGAGAAATGCGAGCTAGATTTTCGCACTCCGCCAGGCTTGA | 2242 |
| Qy | 739 | pArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAl | 759 |
| Db | 2243 | CCGALACGTAAATGCTGAGATTGGAAATAAAGACANTAAATTCAGCAATTTGGAGGAAC | 2302 |
| Qy | 759 | aPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaProAspAsnArgGl | 779 |
| Db | 2303 | TTTTTACATCAGACACTGGCTTGTTCAGGATATATAAAGTGAAGCAGCCTGCACACAGGA | 2362 |
| Qy | 779 | uThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLysTyrLeuSerLy | 799 |
| Db | 2363 | GACACTAGGTCACAAACCTCGNGTCACCAACATCGTCCCAACACAGAGTATTGTCAAA | 2422 |
| Qy | 799 | sLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPheLysLysGlyLy | 819 |
| Db | 2423 | GAAGACTACTAAAGGAGAGCGTGGCTACGTTAAAGATTAAGCTAGTGTTTAAGAAAGGCA | 2482 |
| Qy | 819 | sLysIleSerLysLysThrVal | 826 |
| Db | 2483 | GAAGACCTCTAAGAAGACTGTT | 2504 |
| RESULT 3 | | | |
| BC044321 | | | |
| LOCUS | BC044321 | 4513 bp | mRNA |
| DEFINITION | Xenopus laevis, clone IMAGE:4684177, mRNA. | linear | HTC 23-JAN-2003 |
| ACCESSION | BC044321 | | |
| VERSION | BC044321.1 | GI:27892209 | |
| KEYWORDS | HTC. | | |
| SOURCE | Xenopus laevis (African clawed frog) | | |

| | | | | | |
|------------|---|-------------|------|--------|-----------------|
| BC044321 | BC044321 | 4513 bp | mRNA | linear | HTC 23-JAN-2003 |
| LOCUS | Xenopus laevis, clone IMAGE:4684177, mRNA. | | | | |
| DEFINITION | BC044321 | | | | |
| ACCESSION | BC044321 | | | | |
| VERSION | BC044321.1 | GI:27882209 | | | |
| KEYWORDS | HTC. | | | | |
| SOURCE | Xenopus laevis (African clawed frog) | | | | |
| ORGANISM | Xenopus laevis | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; | | | | |
| | Xenopodinae; Xenopus; Xenopus. | | | | |
| REFERENCE | 1 (bases 1 to 4513) | | | | |
| AUTHORS | Klein,S. and Strausberg,R. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene | | | | |
| | Collection (XGC), National Institute of Child Health and Human | | | | |
| | Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD | | | | |
| | 20892-7510, USA | | | | |
| REMARK | NIH-MGC Project | | | | |
| COMMENT | Contact: XGC help desk | | | | |
| | Email: cgapbs-r@mail.nih.gov | | | | |
| | Tissue Procurement: Dr. Igor David | | | | |

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Brin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 94 Row: 1 Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers

1..4513
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:4684177"
/tissue_type="Embryo, stage 31/32, xenopus"
/clone_lib="NCHD_XGC_Emb4"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4513
Score: 3918.00 Matches: 751
Percent Similarity: 93.38% Conservative: 25
Best Local Similarity: 90.37% Mismatches: 39
Query Match: 89.03% Indels: 17
DB: 3 Gaps: 5

US-10-028-384-2 (1-826) x BC044321 (1-4513)

QY 176 ThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuThrPheAsnGln 195
DB 781 GTATTAGTGGCCCTTACATCAATCTCCACTTCTCTGCTCACTCGAAGCTGTGGAACCCAG 840
QY 196 GlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSer 215
DB 841 GGAGCGGGACCTCTAGCTGCCCTTCATTCATAGTCAGGATACATATCCAGATCA 900
QY 216 ValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyr 235
DB 901 GTAGCTGGATCATTTGACAATGAAGGCTGCTATCTTTGCACCTGCAGTTATATATAT 960
QY 236 LeuTTPValIleValSerValIleThrGlySerValPheThrMetCysCysCysLeuSer 255
DB 961 TTGTGGTAAATCTGTAAACAGGCTCTCTCTGGGCAATAGGTGTGTGTCTCT 1020
QY 256 TyrPheTyrMetValSerAlaTTPGlyTyrValPheIleIleAsnLeuLeuProLeu 275
DB 1021 TATTTTATATGTTTCCGCTCGGGTGTATGTGTATTCATTAATCTTATTCACCTA 1080
QY 276 HisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSer 295
DB 1081 CATGTGTTGTACTGTTACTGATGCAGAGATACAGCAAGAGAGTCTACATAGCTTACAGC 1140
QY 296 ThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGlnPro 315
DB 1141 ACTTTTACATTTGGGTTTAACTACTATCAATGAGATTCCTTTGGGATTTACGCCA 1200
QY 316 IleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuGlnAlaTyrAla 335
DB 1201 ATACGGACAGTGAACATATGGAGCTGCAGGGTCTTTGCCCTGCTGCAAGCTATGCC 1260
QY 336 PheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPheLeu 355
DB 1261 TTCTCTCAGTACTTGCAGAGCAAACTAAGTAAGCAGGATTTCCAGACACTGTTTTCCTC 1320
QY 356 GlyValSerLeuAlaAlaGlyValPheLeuSerValIleTyrLeuThrTyrThrGly 375
DB 1321 GGTGTATGTTGGCAGCTGGAGCTGTTTCTCTACTGTCATCTATTGACTTATACAGGC 1380
QY 376 TyrIleAlaProTTPSerGlyArgPheTyrSerLeuTTPAspThrGlyTyrAlaLysIle 395
DB 1381 TACATGTCTCATGAGTGGAGATTTTACTCATTGGGACACAGGGTATGGAATAATC 1440
QY 396 HisIleProIleIleAlaSerValSerGluHisGlnProThrTyrTTPValSerPhePhe 415
DB 1441 CATATCCCATCATTTGATCTGTCTGAACATCAACCAACATGGGTGCTCTCTCTTT 1500
QY 416 PheAspLeuHisIleLeuValCysThrPheProIleGlyLeuThrPheCysIleLysAsn 435
DB 1501 TTGTATCTGCATATATTAGTATGATATTTCTCTGAGGACTTTGGTTCATCAATAAAT 1560
QY 436 IleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGly 455
DB 1561 ATCAACGATGAAGAGTTTTTGTCTGTATCAATAGTGCCTGTTTACTTTGCTGGC 1620
QY 456 ValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAla 475
DB 1621 GTAATGGTTCGTTGATGCTGACCTGCTCTGCTGTTGTCATGCTGTCGGGCAATGCC 1680
QY 476 PheSerAsnValPheGluHisTyrLeuGlyAspMetLysArgGluAsnProVal 495
DB 1681 TTCTCCATGTTTTTGGACATTTATCTGGGTGATGATATGAAGCGGGAACCCCGCAGTG 1740
QY 496 GluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAlaGly 515
DB 1741 GAAGACAGCAGTGACGAAGAG-----CGAAATTTCTGTGACCTCTATGATAGGCTGGC 1794
QY 516 LysValAlaGlyHisAlaThrGluGlnGluLysThrGluGlyLeuGlyProAsnIle 535
DB 1795 AAAGTGAGAAAGCATGTGTCGAGCAGGAGAGAGAGAGAGGTCTTGGCCCTTAATATA 1854
QY 536 LysSerIleValThrMetLeuMetLeuMetLeuMetPheAlaValHisCysThr 555

```

Db 1855 AAGAGTATAGTACCTATGTTGATGCTATGCTGTTGATGCTGTTGCTGTACACTGTACG 1914
QY 556 TTPValThrSerAsnAlaTyrSerProSerValValLeuAlaSerTyrAsnHisAsp 575
Db 1915 TGGGTAACATGACATGCTTCTCAGCCCTAGTGTGATGCTGCTACATCAGAT 1974
QY 576 GlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnThr 595
Db 1975 GGAACACGAAATATCTTGATGATCTTCAGAGAGCATACTACTGGCTAAGCGCAGATACA 2034
QY 596 AspGluHisAlaAlaTyrMetSerTrpTyrAspTyrGlyTyrGlnIleAlaGlyMetAla 615
Db 2035 GATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2094
QY 616 AsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGly 635
Db 2095 AATCGAACCATCTTATGATAGCAACAATAGCTGGAGCAATAGCCACATAGCTTGGTGGG 2154
QY 636 LysAlaMetSerSerAsnGluThrAlaAlaTyrLysIleMetArgThrLeuAspValAsp 655
Db 2155 AAGGCAATGCTTCAATAGCAACCGCTGCTTATGAAATATGAAAGCTTGGATGATGAT 2214
QY 656 TyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPhe 675
Db 2215 TATGTAATAATAATATTTGGGGGTGTAATGCTGATATCTGCTGATGATCAACAATTC 2274
QY 676 LeuTrpMetValArgIleAlaGlyGlyGlyHisProLysAspIleArgGlnSerAspTyr 695
Db 2275 CTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2334
QY 696 PheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCys 715
Db 2335 TTCA--CCACAGAGAGATTTGCTGTAGACAAAGCTGCTCAACCTTGGCTCAATTC 2392
QY 716 LeuMetTyrLysMetSerTyrArgPheGlyGluMetGlnLeuAspPheArgThrPro 735
Db 2393 CTCATGATATAGATGCTCATATACCGGTTGGTGAATGAGCTTGATTTCCGACACCA 2452
QY 736 ProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHis 755
Db 2453 CTGTGATTTGACCGGACACGTAATGCTGAAATGCGAATAGGATATTAATTAATCAAGCAT 2512
QY 756 LeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaPro 775
Db 2513 TTGGAGAGAGCTTTTACATCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2572
QY 776 AspAsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnLys 795
Db 2573 GAAACACAGAGAGGCACTTGATCATAAAGCTAGAGTAACCAATATTTGACCAACAGAAA 2632
QY 796 TyrLeuSerLysLysThrThrLysArgLysArgGlyTyrIleLysAsnLysLeuValPhe 815
Db 2633 TATTTATCAAAAAGACATCTTAAGAGAGCGTGGCTACATTAGGAACAAGCTGACACTG 2692
QY 816 LysLysGlyLysLysIleSerLysLysThrVal 826
Db 2693 AAGAAAGGCAAGACACCAACAAAGAGTCTGCTG 2725

RESULT 4
LOCUS AY310157
DEFINITION Rattus norvegicus Ac1573 mRNA, complete cds.
ACCESSION AY310157
VERSION AY310157.1 GI:32264636
KEYWORDS HTC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2334)
AUTHORS Xu,C.S., Li,W.Q., Li,Y.C., Han,H.P., Wang,G.P., Chai,L.Q.,

```

```

Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F., Ma,H.,
Wang,L., Wang,S.P., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
Liver regeneration after PH
2 (bases 1 to 2334)
Xu,C.S., Li,W.Q., Li,Y.C., Han,H.P., Wang,G.P., Chai,L.Q.,
Yuan,J.Y., Yang,K.J., Yan,H.M., Chang,C.F., Zhao,L.F., Ma,H.,
Wang,L., Wang,S.P., Shi,J.B., Rahman,S., Wang,Q.N. and Zhang,J.B.
Direct Submission
Submitted (29-MAY-2003) Henan Bioengineering Key Lab, Henan Normal
University, No. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R.
China
Location/Qualifiers
1..2334
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
232..2334
/notes="liver regeneration related protein LRRG109"
/codon_start=1
/product="Ac1573"
/protein_id="AAP78765.1"
/db_xref="GI:32264637"
/translation="MITAGLIHWLNTLNTITVHIRDVCFVLAPTFSGLTSTISFLLTR
ELNQAGLLAACFIATVPGYISVAGSFDNEGIAIFALQFTYVYVHGVVLPFPA
HRSPPGSSGSLTISLVAALPSGALVHARPELYRSVAGVYFIIINLPLHVFV
LLLMORYSKVYIYFALLQAVAFLOVLRDLTKQBFOTLFLGVSLAAGAVFLSVY
LVTGKQYTFOLLHOCATISLHSGGLSLIYFLCVRSQOAYGSAKSLTKVEYLAV
KYRKHVTEQKPEEGIGPNIKSIYVTMLMLMFMENAVHCTWTNSYSSPSVVLASYN
HDGRNLLDDFREAYFWLRONTDEHARVSWMDYGYIAGMANRITLVNNWNNSHI
ALVGKMSNETAAYKIMRSLDYVYVIFGVYGGDDINKFLMMVRIAGEHPKD
IREGYTQGGEPFRVDRAGPTLLNCLMYMSYRFGEMQDFRTPPGDFTRNABIG
NKDIFKHLBEAFTSLHMLVRIYKVPADNRNLTGHKPRVTNIIPKQVLSKKDLAAS
TEERGPHTGLTVLSBEEOKWRLSVGLSENVLPDCQGHSTDSNLHBOEQO
MWTNGLGSEQVSAKGGHSPFCQDRAPFSDRFLNLCPLFSLHWDNSWQNCNVRNT
ESREKQCAQYPCQPRV"
Alignment Scores:
Pred. No.: 1,49e-264 Length: 2334
Score: 2763.00 Matches: 557
Percent Similarity: 74.80% Conservative: 10
Best Local Similarity: 73.48% Mismatches: 27
Query Match: 62.78% Indels: 166
Dbs: 3 Gaps: 7
US-10-028-384-2 (1-826) x AY310157 (1-2334)
QY 106 PheAsnTyrArgSerThrHisHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrp 125
Db 109 TTTAACTATAGTCAACTCATCATCTTCATCTCATGGGTTCTATGAGTCTTAATTCG 168
QY 126 PheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGly 145
Db 169 TTTGATGAAAGAGCATGCTACCCACTAGGAAGAATCTGGGTGGCCACCGCTTATCCACGA 228
QY 146 LeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHis 165
Db 229 TTGATGATACAGCTGGCTTATTCATGCTTTAAATACATGTAACATACAGTTTAC 288
QY 166 IleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThr 185
Db 289 ATCAGAGATGTATGTGTCTTCTTGACCAACTTTTAGCGGCGCTTATCATCTATATATCAG 348
QY 186 PheLeuLeuThrArgGluLeuTrpAsnGlnIleValaGlyLeuLeuAlaCysPheIle 205
Db 349 TTCTGCTACTAGAGAACTTTGGACACGAGGAGGAGGCTCTTAGCTGCTGCTTCAAT 408
QY 206 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIle 225
Db 409 GCTATTGTACCGGGTACATATCTCGGTGCGGTAGCGGATCTCTTTGATACGAAGGCATT 468

```

ORIGIN

```

QY 226 AlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGlySer 245
DB 469 GCGATTTCGCACTTCAGTTCACCTACTTATGGGCCATTCT-----GGGGTT 519
QY 246 ValPheTrpThrMetCys----- 251
DB 520 GTCCTCTG-CTTTCCTGCCCATGCTCTCCAGGCCCTTGGTTCTCAGGGCATTCACT 578
QY 252 -----CysCysLeuSer 255
DB 579 CACCATTCTCTGGTGGCCGCTGCTCTCTAGTGGGGCGCTGGTACATGTTGCTCGGCC 638
QY 256 -TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleLeuAsnLeuIleProLe 275
DB 639 TGAGCTCTCAGGGTCTCTCGTGGGAGGTATGTTTCATCAACCACTCACTCCCTCT 698
QY 275 uHisValPheValLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyrSe 295
DB 699 CCATGTGTTGCTGCTGCTGATGCAGAGGTACAGCAAGAGAGTCTACATA----- 750
QY 295 rThrPheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnPr 315
DB 750 ----- 750
QY 315 oIleArgThrSerGluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAl 335
DB 751 -----GGTGTCTTTGGCTGCTGCGAGGCTTATGC 779
QY 335 aPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnPheGlnThrLeuPhePheLe 355
DB 780 GTTTCGCACTACTGAGAGCGGTGACAAAGCAGAGGTCCAGACCTTCTCTTTT 839
QY 355 uGlyValSerLeuAlaGlyAlaValPheLeuSerValIleTyrLeuThrTyrThrGl 375
DB 840 GGGTGTCTCACTAGCTGAGCGCTGTGTTCTTGTCTATCTATCTGACATAC----- 894
QY 375 yTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaValIle 395
DB 895 -----ACAGG-TATGCAAAAT 910
QY 395 eHisIleProIleAlaSerValSerGluHisGlnProThrTrpValSerPhePh 415
DB 911 ACACATTCCAATATTGTCATCTAGTGTCTGAACATCAGCCTACGACATGGGTCTCTTCTT 970
QY 415 ePheAspLeuHisIleLeuValCysThrPhePheAlaGlyLeuTrpPheCysIleLysAs 435
DB 971 CTTTGACTCATATTCTTGTTGTACCTTCCAGCGAGGCTTATGTTCTGCAATCAAAA 1030
QY 435 nIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAlaGl 455
DB 1031 TATCAACGATGAAAGAGTATT----- 1052
QY 455 yValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAl 475
DB 1052 ----- 1052
QY 475 aPheSerAsnValPheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProProVa 495
DB 1052 ----- 1052
QY 495 lGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTyrAspLysAla-G 515
DB 1053 -----GGCAG 1057
QY 515 lLysValArgLysHisAlaThrGluGlnLysThrGluGluGlyLeuGlyProAsnI 535
DB 1058 GGAAAGTGAGAAAGCATGTGACGGAGCAGGAGGAGCGGAGGAGGGCTTGGGCCCAACA 1117
QY 535 lLysSerIleValThrMetLeuMetLeuMetLeuMetMetPheAlaValHisCysT 555
DB 1118 TCAGAGCATCGTGACCATCTGCTATGCTATGCTCTCTCATGATGTTGCTGTGCACTGCA 1177
QY 555 hrTrpValThrSerAsnAlaTyrSerSerProSerValValLeuAlaSerTyrAsnHisA 575

```

```

DB 1178 CTGGGTCACGACGACCCCTACTCCAGCCCACTGGTGGCTCGGCTCTCAATCATG 1237
QY 575 spGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyrPheTrpLeuArgGlnAsnT 595
DB 1238 ATGGAAACAGAAATATCTTAGATGATTTAGAGAAGCATACTTTTGGTTGAGACAAAACA 1297
QY 595 hrAspGluHisAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetA 615
DB 1298 CGGATGAACACGCCAGGCTCATGCTCGTGGGACTATGCTATCAGATTCTGGGATGG 1357
QY 615 laAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValG 635
DB 1358 CCACACAGGACTACTCTGTAGACATAACACCTCGAACACACAGCCACATTCGCTAG 1417
QY 635 lLysAlaMetSerSerAsnGlnThrAlaAlaTyrLysIleMetArgThrLeuAspValA 655
DB 1418 GAAAAGCTATGTTCTTAAATGAAACAGCCGCTATATAAATCATGAGTCTCTTGTATGTTG 1477
QY 655 spTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysP 675
DB 1478 ATTATGTCTTGTCTATTTTGGAGGAGTTATTGCTATTCTGTGTGATGATATCAACAAGT 1537
QY 675 heLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleArgGluSerAspT 695
DB 1538 TCTCTGATGTCTAGGATAGCTGAAGGGGAGCAGCCCAAGACATTTCGGGAAGGGGACT 1597
QY 695 yTrPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnC 715
DB 1598 ACTTCACCCAGCAGGAGGAGTTCGAGTGGACAGAGCGGGTCTCTCTACTCTCTGTTAAAT 1657
QY 715 yLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrP 735
DB 1658 GCCTTATGTATATAATGTCATACAGATTGGAGAAATGCGACTAGATTTTCGCACAC 1717
QY 735 roProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysH 755
DB 1718 CCCCCGGTITTCACCGAACACGTAATGCTGAGATCGGAAATAGGACATTAAATTTCAAGC 1777
QY 755 lLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLysAlaP 775
DB 1778 ATTTGGAGAAAGCCTTTACATCAGACACTGCTGCTGTGAGATATATAAAGTGAAGCAC 1837
QY 775 roAspAsnArgGluThrLeuAspHisLysProArgValThrAsnIlePheProLysGlnL 795
DB 1838 CCGACACAGGAGGACATTAGTTCACAGCCTCGAGTCACCAACATCTTCCCCAACACA 1897
QY 795 yTyrLeuSerLysLys-----ThrThrLysArgLysArgGlyTyrIleLysA 811
DB 1898 AGTATTTCGTAAGAGGATCTTCAGCATCCACAGAGGAGGAGGAGGAGGAGGAGGAG 1957
QY 811 snLysLeuValPhe-----LysLysGlyLysLysLysLysLysLysLysLysLys 822
DB 1958 GCGTCAGATGCTCAGTGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 2007

RESULT 5
AK030363 3802 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus adult male pituitary gland cDNA, RIKEN full-length
DEFINITION enriched library, clone:530402C17 product:OLIGOSACCHARYL
TRANSPERASE SPT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
homolog (Mus musculus), full insert sequence.
ACCESSION AK030363
VERSION AK030363.1 GI:26326354
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carinici, P. and Havashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)

```



```

MEDLINE
PUBMED
REFERENCE
AUTHORS
Carrincci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Sugahara,Y., Shibata,K.,
Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
MEDLINE
PUBMED
Genome Res. 10 (10), 1617-1630 (2000)
11042159
AUTHORS
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Suna,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
MEDLINE
PUBMED
Genome Res. 10 (11), 1757-1771 (2000)
20530913
REFERENCE
11076861
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
MEDLINE
PUBMED
Nature 409, 685-690 (2001)
5
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
MEDLINE
PUBMED
Nature 420, 563-573 (2002)
6
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hirakawa,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saich,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE
Direct Submission
JOURNAL
MEDLINE
PUBMED
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
FEATURES
source
1. 3802
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:5330402C17"
/db_xref="taxon:10090"
/clone="5330402C17"
/sex="male"
/tissue_type="pituitary gland"

```

```

/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
133. .2250
/note="unnamed protein product; OLIGOSACCHARYL TRANSFERASE
STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
homolog [Mus musculus] (SWISSPROT|P46978, evidence: PASTY,
99.7%ID, 100%length, match=2115)
putative"
/codon_start=1
/protein_id="BAC26921.1"
/db_xref="GI:26326355"
/translation="MTKLGFLRLSYEKQDTLLKLLSMAVLSFSTRPAVLRFESV
LHPEDFYNTRTTFLAEQFYKHNFDRAWYPLGRIIGSTIYGLMTAAAIYHV
LHPEDFYNTRTTFLAEQFYKHNFDRAWYPLGRIIGSTIYGLMTAAAIYHV
SVAGSYNEGTAIFCMLLTYMMIKAVKTSIYMAAKSCALFVPMVSSGVGVFLNL
IPLHLVLMLTGRFSRIYVAVCTVVCGLTILSMQISFVGFOPVLSRMAAGVFGGL
COIHFADVYLCSLNPOEPLFRSVISLVGFLLTVGALLMTGKISPTGRFYSLL
DPSYKNNIILIASVSEHQPTTASSYFYLQLLVFMPVGLYCCSNLSLDAKIFIMY
GVTSNYSFSAVMRLMLVAFVCMILSGISGVSVQVSTYMNLDI SRPKKSKQDQSTY
PIKNEVASGMILVNAFFLITYTFHTSWTSEAYSSPSIVLSARGDGSRIIFDREA
YYLRLHNTPEDAKVMWDYQITAMANTILVDNNTWNTNTHSRVGOAMSTEKA
YEIMRELDVSVLVIFGLTYGSDDDINKPLMVRIGGSTETGRHKENDYTPGEGF
RVDRGSPVLLNCLMYKVCYRFQGVTEAKRPPGDFVRNAEIGNKDFELVLEAY
TTEHVLVRIYKVDLDRGLSRT"
ORIGIN
Alignment Scores:
Pred. No.: 7,41e-214 Length: 3802
Score: 2259.50 Matches: 435
Percent Similarity: 70.46% Conservative: 104
Best Local Similarity: 56.86% Mismatches: 155
Query Match: 51.34% Indels: 71
DB: 3 Gaps: 8
US-10-028-384-2 (1-826) x AK030363 (1-3802)
QY 58 LeuSerGlyGlyLeuSerGlnProAla---GlyTrpGln----- 69
Db 19 CTGCTGTGAGGGTTGAGTGCGCGCTGAACGGATGGCAGCGGAGCAGAGTGGGTTCTG 78
QY 69 ----- 69
Db 79 AGGAGCATCGCTGAGGTATTGTAATATCATCATCTGCCACCCATTGATGTCAGATGACT 138
QY 70 -----SerLeuLeuSerPheThrIle 76
Db 139 AAGCTTGAGTATTTTGGCGATGTTCTCTATGACAGCAGCAGACACTTCTAAAGCTTCTCATC 198
QY 77 LeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGlu 96
Db 199 CTGTCGATGGCTGCTGTGTATCTTTTCTACTCGCTCTTTTCTGTGTGTGAGATTGAA 258
QY 97 SerIleIleHisGluPheAspProTrpPheAsnTrpArgSerThrHisLeuAlaSer 116
Db 259 AGTGATCATCATGAGTTTGTATCGGTACTTTTAAATATCGGACTACCCGGTTTCTGGCTGAG 318
QY 117 HisGlyPheTrpGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArg 136
Db 319 GAGGGTTTATAAATTCATATCCATATCGTTTGTAGACCGGCTGGTACCCTTTGGGCCGA 378
QY 137 IleValGlyGlyThrValTrpGlyLeuValIleThrAlaGlyLeuIleHisTrpIle 156
Db 379 ATCATTTGGAGGACAAATTACCAGAGTTTAAATCATCATTCTGTCTGCAATCTACCATGTA 438
QY 157 LeuAsnThrLeuAsnIleThrValHisIleAtqAspValCysValPheLeuAlaProThr 176
Db 439 CTCCATTCTTCCATATCACCATTACATTCCGGAATGTCTGTGTTTCTTCTGGCCCGCTT 498
QY 177 PheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGly 196
Db 499 TTCTCTCTTTTCCACCACCATCTGTTACGTACCACTTACCACAAAGAGCTCAAGCATGCAGCA 558
QY 197 AlaGlyLeuLeuAlaAlaCysPheIleAlaValProGlyTrpIleSerArgSerVal 216

```


| | | |
|------|---|------|
| 559 | GCTGGGCTTCTCCTGCCATGATTGCTAGTTCTCGGTATATTTCTCGACTGTA | 618 |
| 217 | AlaGlySerPheAspAenGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrrLeu | 236 |
| 619 | GCTGCCTCCATATATATGAGGAATGCTATCTTTTCATGCTGCTCTTACTACATG | 678 |
| 237 | TripValLysSerValLysThrGlySerValPheThrThrMetCysCysLeuSerTyrr | 256 |
| 679 | TGGATCAAGCATGAGAAGCTGGTTCATCTATTGGGCTGCCAAGTGCGCCGCTTAT | 738 |
| 257 | PheTyrrMetValSerAlaTrpGlyGlyrValPheIleIleAenLeulleProLeuHis | 276 |
| 739 | TTCTACATGGCTCTTCATGGGAGGCTATGTTCTTGATCAACTTGATTCCTCTACAT | 798 |
| 277 | ValPheValLeuLeuLeuMetGlnArgTyrrSerLysArgValTyrrIleAlaTyrrSerThr | 296 |
| 799 | GTCCTGGTGTAAATGTCACAGCCGTTTTCTCACGGATCTACGTAGGCTACTGTACT | 858 |
| 297 | PheTyrrileValGlyLeuLeuleuSerMetGlnIleProPheValGlyPheGlnProIle | 316 |
| 859 | GTITACTGCCTGGGACCATTCTTCTATGACAGATTCTCTTGTTGGTTCAGCCGCTC | 918 |
| 317 | ArgThrSerdLuisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrrAlaPhe | 336 |
| 919 | CTTTTCATCAACAACATGCGAGCTTTGGAGTGTGTGCTCTCTGCAGATCCCATGCTTC | 978 |
| 337 | LeuGlnTyrrLeuArgAspArgLeuthrLysGlnGluPheGlnThrLeuPhePheLeuGly | 356 |
| 979 | GTAGATTACCTGTGCAGCAAGTTGAATCCACAGCAATTCGAAGTTC-TTTTCCGAGTGT | 1038 |
| 357 | ValSerLeuAlaAlaGlyAlaValPheLeuSerVal---IleTyrrLeuThrTyrrThrGly | 375 |
| 1039 | AITCTCCCTG--GTGCTTTGCTCCTCATCTGGGAGCTCTCCATCGTCAACHAGGA | 1095 |
| 376 | TyrrileAlaProTrpSerGlyArgPheTyrrSerLeutrpAspThrGlyTyrrAlaLysile | 395 |
| 1096 | AAAAATTTCTCCCTGGACAGGGCTTTCTACTCTCTGCTGGATCCTCTTATGTAAGAAT | 1155 |
| 396 | HistLeuProIleIleAlaSerValserGluHisGlnProThrThrTrpValSerPhePhe | 415 |
| 1156 | ARCATTCCTAATTTGCAICTGTTCTGACACACAGCCCAACTGGTCTTCTACTAT | 1215 |
| 416 | PheAspLeuHisLeuValCysThrPheProAlaGlyLeutrpPheCysilleLysAen | 435 |
| 1216 | TTTGATCTACAGCTCCTTGTTCTTCATGTTTCAGTGGCTCTCTATTAAGTTCAGAAC | 1275 |
| 436 | IleAenAspGluArgValPheValAlaLeuTyrrAlaIleSerAlaValTyrrPheAlaGly | 455 |
| 1276 | CTGTCTGATGCTGGATTTTATCATCATGATGTTGACACAGCAGTACTTTTCAGCT | 1335 |
| 456 | ValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAla | 475 |
| 1336 | GTAATGTGGCTCTAATGCTGGPATTTGGCACCTGTATGTGCATTTCTTTCTGGCATGGT | 1395 |
| 476 | PheSerAenValPheGluHisTyrrLeuGlyAsp---AspMetLysArgGluAenProPro | 494 |
| 1396 | GTITCCCAGTGGTGTCCCATATATAGAAAATCTGCACATAAGTCGCCCA----- | 1446 |
| 495 | ValGluAspSerSeraspGluAspAspLysArgAenGlnGlyAenLeutyrrAspLysAla | 514 |
| 1447 | -----GACAGAAG1455 | |
| 515 | GlyLysValArgLysHisAlaThrGluGlnGluLysThrGluGluGlyLeuGlyProAen | 534 |
| 1456 | AGCAAGAAGCAACAGGATTCTACTTACCCTATTAGAAATGAG----- | 1497 |
| 535 | IleLysSerIleValThrMetLeuMetLeuMetLeuMetLeuMetMetPheAlaValHisCys | 554 |
| 1498 | GTGGCGAGTGGATGATACGTGGTCATGCGCTTCTTCTCATCACCCTACACGTTTCATCG | 1557 |
| 555 | ThrirpValThrSerAenAlaTyrrSerSerProSerValValLeuAlaSerTyrrAenHis | 574 |

| | | |
|-----------|---|-----------------|
| 1558 | ACTTGGGTGACAGTGAAGCCATTCTTCTCCCTCCATTTGTTTGTCTGCTGCTGGGG | 1617 |
| 575 | AspGlyThrArgAsnIleLeuAspPheArgGluAlaTyrPheTrpLeuArgGlnAsn | 594 |
| 1618 | GATGGCATAGGATCATTTTTCATGACTCCGAGAAGCGTATTATTGGCTCCGTCACAAT | 1677 |
| 595 | ThrAspGluHisIslaArgValMetSerTrpTrpAspTyrGlyTrpGlnIleAlaGlyMet | 614 |
| 1678 | ACTCCAGAGGATCAAAAGTCATGTCACTGGTGGGATTTATGCTACCCAAATTCTCGAATG | 1737 |
| 615 | AlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVal | 634 |
| 1738 | GCAAAATCCGACAAATTTTATGCGACAAATAACACATGGATATATATACCCATATTTCTCGAGTA | 1797 |
| 635 | GlyIlyValaMetSerSerAsnGlnThrAlaAlaTyrIlyValMetArgThrLeuAspVal | 654 |
| 1798 | GGCGAGCGCATGTCATCCACAGAGAAAAGCCATATGAAATCATGAGGGAGCTTGATGTC | 1857 |
| 655 | AspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLys | 674 |
| 1858 | AGCTATATGCTTCTCATTTTTCGAGCGCTTACTGCGTATTCTTCGGATGATATCAACAAG | 1917 |
| 675 | PheLeuTrpMetValArgIle--AlaGluGlyGluHisProLysAspIleAsnGluSer | 693 |
| 1918 | TTTCTTTGGATGTCTCGGATTTGGAGAGACACAGACAGACACATATAGGAGAT | 1977 |
| 694 | AspTyrPheThrProGlnGlyIlePheArgValAspIlyAsnLysArgProThrLeuLeu | 713 |
| 1978 | GACTACTATATCTCTACTCGGAAATTCGGTGTGTATCGTGAGGGTCTCCGGTGTGCTC | 2037 |
| 714 | AsnCyValLeuMetTyrIlyValMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArg | 733 |
| 2038 | AACTCGCTTATGTACAAATATGTCTACTACCGCTTGGCAGGTCTACACAGAGCCAAAG | 2097 |
| 734 | ThrProGlyPheAspArgThrArgAsnIleGluIleGlyAsnLysAspIleLysPhe | 753 |
| 2098 | CGTCCACAGCGCTTTCACCGCTGTCGAAATGCTGAGATGCGTAAATAAGACTTTTGAGCTT | 2157 |
| 754 | LysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArgIleTyrLysValLys | 773 |
| 2158 | GATGCTCTGGAGGAGCGTATACCACAGACACTGGCTAGTCAGATATATCAAGGTAAG | 2217 |
| 774 | AlaProAspAsnArg | 778 |
| 2218 | GACCTGGATATCGA | 2232 |
| BC028897 | 2656 bp mRNA linear | HTC 04-AUG-2004 |
| BC028897 | Mus musculus cDNA clone IMAGE:4168013, containing frame-shift errors. | |
| ACCESSION | BC028897.1 | GI:22137711 |
| VERSION | BC028897 | |
| KEYWORDS | HTC | |
| SOURCE | Mus musculus (house mouse) | |
| ORGANISM | Mus musculus | |
| REFERENCE | 1. (bases 1 to 2656) | |
| AUTHORS | Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Sherman, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marushina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Bosak, S.A., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Sosak, S.A., McEwan, P.J., McKernan, K.J., Matek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S.G., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villaon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Halton, E., Kettman, M., Madan, A.C., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., | |

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalins, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
 12477932
 2 (bases 1 to 2656)
 DIRECTOR MGC PROJECT
 DIRECT SUBMISSION
 Submitted (01-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: angbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
 Kowis, C.R., Sheed, A.J., Martin, R.G., Muzny, D.M., Manavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 24 Row: 1 Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 31560679
 This clone has the following problem: frame shifted.

FEATURES

source

1. .2656
 /organism="Mus musculus"
 /mol_type="cDNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4168013"
 /tissue_type="Salivary gland, 10 week old female mouse"
 /clone_lib="NCI CGAP_SG2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:

Pred. No.: 5,266-214 Length: 2656
 Score: 2258.50 Matches: 428
 Percent Similarity: 74.44% Conservative: 102
 Best Local Similarity: 60.11% Mismatches: 152
 Query Match: 51.32% Indels: 30
 DB: 3 Gaps: 5

US-10-028-384-2 (1-826) x BC028897 (1-2656)

Qy 70 SerLeuLeuSerPheThrIleLeuAlaTrpLeuAlaGlyPheSerArgLeu 89
 Db 65 ACACTTCTAAAGCTTCATCTCTCGATGCTGCTGCTATCTTTCTACTCGTCTT 124
 Qy 90 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 109
 Db 125 TTGCTGTGCTGAGATTTGAAGTGTCAATCCAGTTGATCGTACTTTAATATCGG 184
 Qy 110 SerThrHisIleLeuAlaSerHisGlyPheThrGluPheLeuAsnTrpPheAspGluArg 129
 Db 185 ACTACCGCGTTCTGCTGAGGAGGGGTTTATAAATTCATATCCATACTGTTGATGACCGG 244
 Qy 130 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 149

Db 245 GCTTGGTACCCTTTGGGCGGAATCAITGGAGGAAACAATTTTACCAGGTTTAAATGATCACT 304
 Qy 150 AlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrValHisIleArgAspVal 169
 Db 305 TCTGCTGCATCTACCATGACTCCATTTCTTCATATACATATTGACATTCGGAATGTC 364
 Qy 170 CysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThr 189
 Db 365 TGTGTTTTCTTGGGCCCACTTTTCTCTCTTTCCACCACCATCGTTACGACCACTTACC 424
 Qy 190 ArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaAlaCysPheIleAlaIleValPro 209
 Db 425 AAAGAGCTCAAGGATGAGGAGCTGGGCTTCTTGTCTGCTGCCATGATTGCTGTAGTTCT 484
 Qy 210 GlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAla 229
 Db 485 GGGTATATTTCTCGATCTGAGTGGCTCTATGATAAAGAGGAATTCATCTTTTTCG 544
 Qy 230 LeuGlnPheThrTyrTyrLeuTrpValIleSerValIleSerThrGlySerValPheTrpThr 249
 Db 545 ATGCTGCTTACTTACTACATGTGGATCAAGGAGTGAAGACTGGTTCATCTATTGGGCT 604
 Qy 250 MetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 269
 Db 605 GCCAAGTGTGCCCTGCTGCTTATTTCTACATGCTCTTCATGGGAGGCTATGTTCTCTG 664
 Qy 270 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArgTyrSerLysArg 289
 Db 665 ATCAACTGATTCCTCTACATGCTGCTGCTAATGCTGACAGCGCGTTTCTCTACCGG 724
 Qy 290 ValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleSerMetGlnIlePro 309
 Db 725 ATCTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 784
 Qy 310 PheValGlyPheGlnProIleArgThrSerGluHisValAlaAlaAlaGlyValPheAla 329
 Db 785 TTGTTGGTTTCCAGCGCGTCTTTCATCAGAACACATGCGACGCTTTGGAGTGTGTG 844
 Qy 330 LeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPhe 349
 Db 845 CTCCTGTCAGATCCATGCTTTCTGATGATTACCTGCGGAGCAAGTGAATCCAGCAATTC 904
 Qy 350 GlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerVal--- 368
 Db 905 GAAGTCTCTTTCCGGAGTGTATCTCCCTG---GTTGGCTTTGCTCTCTCTCTCTCTG 961
 Qy 369 IleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrp 388
 Db 962 GCTCTCTCATGCTAACAGGAAATTTCTCCCTGCGAGGCGTTTCTACTCTCTGCTG 1021
 Qy 389 AspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnPro 408
 Db 1022 GATCCCTCTTATGCTAAGAATAACATTCCTCATTTATTCATCTGTTTCTGAGCACCAGCC 1081
 Qy 409 ThrThrTrpValSerPhePheAspLeuHisIleValCysThrPheProAlaGly 428
 Db 1082 ACAACCTGGTCTTCTACTATTGATCTACACGCTCTGCTGCTCTGCTGCTGCTGCTGCTG 1141
 Qy 429 LeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIle 448
 Db 1142 CTCTATTACTGCTTTAGCAACCTGCTGATGCTCGGATTTTATCATCATGATGATGCTG 1201
 Qy 449 SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal 468
 Db 1202 ACCAGCATGACTTTTTCAGCTGTAATGCTGCGTCTAATGCTGATTTGGCACCCTGTTATG 1261
 Qy 469 CysMetLeuSerAlaIleAlaPheSerAsnValPheGlu-HisTyrLeuGlyAspAspMe 488
 Db 1262 TGCATTTCTTCTGCAATGCTGTTTCTCCAGGCTGCTGCCACATATGAAAACTGACAT 1321
 Qy 488 tLysArgGluAsnProProValGluAspSerSerAspGluAspLysArgAsnGlnG1 508
 Db 1322 AAGTCGCCCA----- 1331

```

QY 508 yAsnLeuTyrAspIysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThrCl 528
DB 1332 -----GACAAAGAGACCAAGACCAAGGATTCCTACTTACCCTATTAAAGAAATGA 1381
QY 528 uGluGlyLeuGlyProAsnIleLysSerIleValThrMetLeuMetLeuMetLeuMetLeuMe 548
DB 1382 G-----GAGCGAGTGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1423
QY 548 tMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValVa 568
DB 1424 CACCTACAGCTTTCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1483
QY 568 lLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyr 588
DB 1484 ACTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1543
QY 588 rPheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrCl 608
DB 1544 TTATTGGCTCCGTCACAACTCCAGAGGATGCAAAAGTCATGTCATGTCGATGATGATG 1603
QY 608 YTrpGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAs 628
DB 1604 CTACCAAAATTACTGCAATGCGCAATTCGACATTTTAGTCGACAAATAACACATCGAATAA 1663
QY 628 nSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGluThrAlaAlaTyrLysI 648
DB 1664 TACCATATTTCTCGATGAGGCGAGCAATGGCATCCACAGAAAGAAAGCCCTTGAAT 1723
QY 648 eMetArgThrLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSe 668
DB 1724 CATGAGGAGCTGATGTCAGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 1783
QY 668 rGlyAspAspIleAsnLysPheLeuTyrMetValArgIle--AlaGluGlyGluHisPr 687
DB 1784 TTCGATATATCAACAAGTTCTTTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1843
QY 687 oLysAspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAl 707
DB 1844 AAGACACATTAAGGAGATGACTACTACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1903
QY 707 aGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyG 727
DB 1904 GGGTCTCGGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1963
QY 727 uMetGlnLeuAspPheArgThrProGlyPheAspArgThrArgAsnAlaGluIleG 747
DB 1964 GGTCTACACAGAACCAAGCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2023
QY 747 yAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuVa 767
DB 2024 TAATAAGACCTTTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2083
QY 767 lArgIleTyrLysValLysAlaProAspAsnArg 778
DB 2084 CAGGATATACAGGTAAAGGACCTCGGATAATCGA 2117

RESULT 7
AK081547 2730 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
DEFINITION library, clone:G130038C18, product:OLIGOSACCHARYL TRANSFERASE STT3
SUBUNIT HOMOLOG (BS) (INTERGAL MEMBRANE PROTEIN 1) homolog [Mus
musculus], full insert sequence.
ACCESSION AK081547
VERSION AK081547.1 GI:26100006
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 Carninci, P. and Hayashizaki, Y.
AUTHORS

```

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS

High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Kashiwagi, K.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Katsuhiko, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, Y.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2730)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saio, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN) Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
Location/Qualifiers
1. 2730
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:C130038C18"
/db_xref="taxon:10090"
/clone="C130038C18"

/tissue type="head"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="16 days embryo"
 125..2243

misc_feature

/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
 (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
 (SWISSPROT P46978, evidence: FASTY, 99.7%ID, 100%length,
 match=2115)
 putative"
 2717..2722

polyA_signal

/note="putative"
 2730

polyA_site

/note="putative"

ORIGIN

Alignment Scores:

Pred. No.: 1.1e-213 Length: 2730
 Score: 2255.50 Matches: 436
 Percent Similarity: 70.50% Conservative: 104
 Best Local Similarity: 56.92% Mismatches: 154
 Query Match: 51.25% Indels: 72
 DB: 3 Gaps: 8

US-10-028-384-2 (1-826) x AK081547 (1-2730)

QY 58 LeuSerGlyGlyLeuSerGlnProAla---GlyTrpGln----- 69
 DB 11 CTGCTGTCAAGGTTGAGTGGCCCGCTGAACGGATGGCAGGGGAGCAGAGTGGTTCCTG 70
 QY 69 ----- 69
 DB 71 AGGAGCATCCGTGAGGTATTTGAATATCATCATGTTGCCACCCATTGATCAAGATGACT 130
 QY 70 -----SerLeuLeuSerPheThrIle 76
 DB 131 AGCTTGGATTTTCGGATGTCCTATGAGAGCAGACACATCTCAAGCTTCTCATC 130
 QY 77 LeuPheLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGlu 96
 DB 191 CTGCGATGGCTGCTGCTGCTATCTTTTCTACTGCTCTTTTGTGCTGCTGAGATTGAA 250
 QY 97 SerIleIleHisGluPheAspProTrpPheAsnTrpArgSerThrHisHisLeuAlaSer 116
 DB 251 AGTGTCATCCATGATTTGATCCGACCTTTAATATATGACTACCCCGTTCTGGCTGAG 310
 QY 117 HisGlyPheTrpGluPheLeuAsnTrpPheAspGluArgAlaTrpTrpProLeuGlyArg 136
 DB 311 GAGGGGTTTATAAAATCCATAACTGGTTTGTATGACCGGGCTTGGTACCCCTTTGGGCCGA 370
 QY 137 IleValGlyGlyThrValTrpProGlyLeuMetIleThrAlaGlyLeuIleHisTrpIle 156
 DB 371 ATCATTTGAGGAGAACATTTACCGAGTTTAAATGATCACTTCGTGCAATCTACCATGTA 430
 QY 157 LeuAsnThrLeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThr 176
 DB 431 CTCATTTCTTCCATATCACCATTGACATTCGGAAATGCTGTGTTTCTTGGCCCGCTT 490
 QY 177 PheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGly 196
 DB 491 TTCTCTCTTTCACCCATCGTTACGPACCACTTACCAAGAGCTCAAGAGTGCAGGA 550
 QY 197 AlaGlyLeuLeuAlaAlaCysPheIleAlaIleValProGlyTrpIleSerArgSerVal 216
 DB 551 GCTGGGCTTCTTGTCTGCTGCCATGATTCGTAGTTCTGCGGTATATTTCTCGATCTGTA 610
 QY 217 AlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTrpTrpLeu 236
 DB 611 GCTGCTCTATGATATGAAGAAATTCATATCTTTTGCATGCTGCTTACTTACTACAT 670
 QY 237 TrpValIleSerValIleThrGlySerValPheTrpThrMetCysCysLeuSerTrp 256
 DB 671 TGGATCAAGGCAGTGAAGACTGGTTTCCATCTATTGGGCTGCCAAGTGTGCCCTTCGCTTAT 730

QY 257 PheTrpMetValSerAlaTrpGlyGlyTrpValPheIleLeuLeuLeuPheLeuHis 276
 DB 731 TTCATATGGTCTCTTATGAGGAGGCTATGTGTCTCTGATCACTTGTATCTCTACAT 790
 QY 277 ValPheValLeuLeuLeuMetGlnArgTrpSerIleValTrpValTrpValTrpValTrp 296
 DB 791 GTCTCTGGTGAATGCTGACAGGCGTCTTCTCACCGATCTACGTAGCTACTGCTACT 850
 QY 297 PheTrpIleValGlyLeuLeuLeuSerMetGlnIleProPheValGlyPheGlnProIle 316
 DB 851 GTTACTCTGCTGGGACCAATCTTCTATGAGATTTCTTGTGTGTTTCCAGCCCGCT 910
 QY 317 ArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTrpAlaPhe 336
 DB 911 CTTTCATCAGAACACATGGCAGCCCTTTGGAGTGTGTTCTCTCTCTCAGATCCATGCTTC 970
 QY 337 LeuGlnTrpLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGly 356
 DB 971 GTAGATTACCTGGCAGCAAGTTGAATCCACAGCAATTCAGAGTTCTTTTCCGAGTGT 1030
 QY 357 ValSerLeuAlaAlaGlyAlaValPheLeuSerVal---IleTrpLeuThrTrpValSerPhe 375
 DB 1031 ATCTCCCTG---GTTGGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1087
 QY 376 TyrIleAlaProTrpSer---GlyArgPheTrpSerLeuTrpAspThrGlyTrpAlaLysIle 395
 DB 1088 AAATTTCTCCCTGGCAGGCGCTTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1147
 QY 395 eHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpValSerPhePhe 415
 DB 1148 TAACATTTCCATTTATGTCATCTGTTTCCAGTGGCTCTATATCTCTCTCTCTCTCTCTCT 1207
 QY 415 ePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleLysAs 435
 DB 1208 TTTTGAICTACGCT 1267
 QY 435 nIleAsnAspGluArgValPheValAlaLeuTrpAlaIleSerAlaValTrpPheAlaG 455
 DB 1268 CTCTCTCATGCTCGGATTTTATCATCATGATGTTGTTGACCAACCTGTTCTCTCTCTCTCT 1327
 QY 455 ValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaIleAl 475
 DB 1328 TGTATGTTGCTGCTTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1387
 QY 475 aPheSerAsnValPheGluHisTrpLeuGlyAsp---AspMetLysArgGluAsnProPr 494
 DB 1388 TGTTTCCCGAGTGTGTCACATATATCAAAATCTTGGACATAAGTCGCCCA----- 1439
 QY 494 oValGluAspSerSerAspGluAspAspLysArgAsnGlnGlyAsnLeuTrpAspLysAl 514
 DB 1440 -----GACAGAA 1447
 QY 514 aGlyLysValArgLysHisAlaThrGluGlnGlyLysThrGluGluGlyLeuGlyProAs 534
 DB 1448 GAGCAAGAAAGCAACAGGATCTTACTTACCTTATTAAGATGAG----- 1490
 QY 534 nIleLysSerIleValThrMetLeuMetLeuMetLeuMetLeuMetPheAlaValHisCy 554
 DB 1491 -GTGCGAGTGGGATGATGATGTCATGTCGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1549
 QY 554 sThrTrpValThrSerAsnAlaTrpSerProSerValValLeuAlaSerTrpAsnHis 574
 DB 1550 GACTTGGGTGACCAAGTGAAGCTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1609
 QY 574 sAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTrpPheTrpLeuArgGlnAs 594
 DB 1610 GGATGGCAGTGAAGATCAATTTTGTATGACTTCCGAGAGCGTATATTGGTCTCGTCACAA 1669
 QY 594 nThrAspGluHisAlaArgValMetSerTrpTrpAspTrpGlyTrpGlnIleAlaGlyMe 614
 DB 1670 TACTCCAGAGGATGCAAAAGTCAATGTCATGGTGGGATATTGGCTACCAAAATTAATCTCAAT 1729
 QY 614 tAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVa 634

```

Db      1730  GCGAAATCGGCAATTTAGTGACATAACATGGAATAATACCATATTCCTCGAGT 1789
Qy      634  LGLYSLAAlaMetSerSerAnGluThraLaalATyLysIleMetArgThrLeuAspVa 654
Db      1790  AGGCGAGCAATGCGATCCACAGAGAAAAGCCTATGAATCATGAGGAGCTTGATGT 1849
Qy      654  lAspTyValLeuValIlePheGlyGlyValIleGlyTySerGlyAspAspIleAsnIly 674
Db      1850  CAGCTATGTGCTGCTATTTTGGAGCCCTTACTGGTATTCCTCGATGATATCAACAA 1909
Qy      674  sPheLeuTrpMetValArgile---AlaGluGlyGluHISpOLysAspIleArgGluLe 693
Db      1910  GTTCTTTGGATGCTCCGATGGAGGAACACAGACAGAGAGACACATTAAAGAGAA 1969
Qy      693  rAspTyPheThrProGlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLe 713
Db      1970  TGACTACTATACTCTTACTGGGAATTCCTGTTGATCGTGGAGGTCTCCGGTCTGCT 2029
Qy      713  uAsnCysLeuMetTyLysMetSerTyTyArgPheGlyGluMetGlnLeuAspPheAr 733
Db      2030  CAACGTGCTTATGTAACAAATGTTACTACCGCTTTGGCAGGCTACACAGAACCCAA 2089
Qy      733  gThrProGlyPheAspArgThrArgAsnAlaGluIleGlyAsnLysAspIleGlyPh 753
Db      2090  CGCTCCACCCAGGCTTACCGGTGTCGAAATGCTGAGATTGGTAATAAGACTTTGAGCT 2149
Qy      753  eLysHISLeuGluGluAlaPheThrSerGluHISrTpLeuValArgIleTyLysVally 773
Db      2150  TGATGTCTCTGGAGGAGCGTATACACAGAACACTGGTAGTCAGGATATACAGGTAAA 2209
Qy      773  sAlaPheProAspAsnArg 778
Db      2210  GGACTGTGATAATCGA 2225

RESULT 8
AK077877
LOCUS
DEFINITION
Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
enriched library, clone:6030401L06 product:OLIGOSACCHARYL
TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
homolog [Mus musculus], full insert sequence.
ACCESSION
AK077877
VERSION
AK077877.1 GI:26097543
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
REFERENCE
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
REFERENCE
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kutsuna, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer

```

```

Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
REFERENCE
4
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
(bases 1 to 3643)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hasegaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Iishi, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akaira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Sushiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://phantom.gsc.riken.jp/
FEATURES
Location/Qualifiers
1..3643
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="PANTOM_DB:6030401L06"
/db_xref="taxon:10090"
/clone="6030401L06"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
93..2208
misc_feature
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
(SWISSPROT|P46978, evidence: FASTA, 99.7%ID, 100%length,
match=2115)
putative"
ORIGIN
Alignment Scores:
Pred. No.: 9,79e-212 Length: 3643
Score: 2238.00 Matches: 427
Percent Similarity: 74.58% Conservative: 104
Best Local Similarity: 59.97% Mismatches: 151
Query Match: 50.85% Indels: 31
DB: 3 Gaps: 6
US-10-028-384-2 (1-826) x AK077877 (1-3643)

```



```

Db      1510 GACITCCGAGAGCATATATGGCTTCGTATATCTCCAGAGGATGCGAAGGTCATG 1569
QY      603 SerTtpTAspTyrGlyTyrGlnIleAlaGlyMetAlaSenArgThrThrLeuValAsp 622
Db      1570 TCTGGTGGATTATGGCTATCAGATACACGTAATGCGAACCAGCAATTTTAGTGGAC 1629
QY      623 AnAnThrTtpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAnGlu 642
Db      1630 AATAACACATGGAATTAACCATATTTCTCGATAGGCGAGCAATGGCGTCCACAGAG 1689
QY      643 ThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGly 662
Db      1690 GAAAAGCCTATGACATCATGAGGAGCTCATGTCAGCTATGCTGCTGTCATTTTGA 1749
QY      663 GlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArgIleAla 682
Db      1750 GGCCTCACCTGGTATCTCTCGATGATATCAACAGTTCTTGGATGGTCCCGATGGA 1809
QY      683 -----GluGlyGluHisProLysAspIleArgGluSerAspTyrPheThrPro 698
Db      1810 GCGAGCACAGATACAGGCAACAT-----ATCAAGGAGATGACTATTATCTCCA 1860
QY      699 GlnGlyGluPheArgValAspLysAlaGlySerProThrLeuLeuAsnCysLeuMetTyr 718
Db      1861 ACTGGGGAGTTCGGTGTGGACCGTGAAGTTCTCCAGTGTCTCAACTGCCCTCATGTAC 1920
QY      719 LysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProProGlyPhe 738
Db      1921 AAGATGTGTACTATCGTCTGGACAGCTTTACACAGANNNNAGCGTCTCCAGGCTTT 1980
QY      739 AspArgThrArgAsnAlaGlnIleGlyAsnLysAspIleLysPheLysHisLeuGlu 758
Db      1981 GACCGTGTCCGAAATGCTGAGATGGGAATTAAGACTTTGATGTGATCCCTGGAGGAA 2040
QY      759 AlaPheThrSerGluHisThrPheLeuValArgIleTyrLysValLysAlaProAspAsnArg 778
Db      2041 GCATATACCACAGACAATGGCTGCTCAGGATATACAAAGGTAAAGGACCTGGATATCGA 2100

RESULT 10
AK087470
LOCUS
DEFINITION
Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus
musculus], full insert sequence.
ACCESSION
AK087470
VERSION
AK087470.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2
Genome Res. 10 (10), 1617-1630 (2000)
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,

```

```

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530313
PUBMED
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
(bases 1 to 2727)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hata, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Kakuwa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
CNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
location/Qualifiers
1. .2727
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FANTOM DB:E130302E03"
/db_xref="taxon:10090"
/clone="E130302E03"
/tissue_type="eyeball"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
108. .2233
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
(SWISSPROT|P46978, evidence: FASTA, 99.7%ID, 100%length,
match=2115)
Putative"
2708. .2713
/note="putative"
2727
/note="putative"
misc_feature
polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.:
Score:
2.47e-199
2112.00
Length:
Matches:
2727
419

```

```

Percent Similarity: 68.49%      Conservative: 107
Best Local Similarity: 54.56%    Mismatches: 164
Query Match: 47.99%             Indels: 79
DB: 3                            Gaps: 8

US-10-028-384-2 (1-826) x AK087470 (1-2727)

Qy 61 GlyLeuSerGlnProAla---GlyTyrGln----- 69
Db 3 GGGTTGAGTGGCCGCTGACGGATGGCAGGGGGACAGAGTGGGTCTCTGAGGAGCATC 62
Qy 69 ----- 69
Db 63 CGTGAGGTATTGGAATATCATCATAGTTGCCACCCATGATGTCAAGATGACTAAGCTTGGGA 122
Qy 70 -----SerLeuLeuSerPheThrIleLeuPheLeu 79
Db 123 TTTTTCGGATTGCTATGAGAGCAGGACACACTTCTAAAGCTTCTCATCTGTCGATG 182
Qy 80 AlaTrpLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheIleuSerIlele 99
Db 183 GCTCCTGTGTATCTTTTCTACTCTGCTTTTTCGCTGCTGAGATTTGAAAGTGTATC 242
Qy 100 HisGluPheAspProTrpPheAsnTyrArgSerThrHisIleAlaSerHisGlyPhe 119
Db 243 CATGAGTTTGATCGGTACTTTAATTATCGGACTACCCGGTTCTGGCTGAGGAGGGTTT 302
Qy 120 TyrGluPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGly 139
Db 303 TATAAATTCATAACTCGTTTGTATGACCGGGCTGTGTACCTTTGGCGCGATGATGGA 362
Qy 140 GlyThrValTyrProGlyLeuMetIleThrAlaGlyLeuIleHisIleTrpIleLeuAsnThr 159
Db 363 GGAACAATTACCCAGGTTTAATGATCACTTCTGCTGCAATCTACCATGACTCTCATTTT 422
Qy 160 LeuAsnIleThrValHisIleArgAspValCysValPheLeuAlaProThrPheSerGly 179
Db 423 TTCATATACCAATTCATTCGAAATGCTGTGTGTTTCTCGCCCGCTTTCTCCCTCT 482
Qy 180 LeuThrSerIleSerThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyValAlaGlyLeu 199
Db 483 TTCACCACTCGTTAGTACCCACTTACCAAGAGCTCAAGATGCGAGGCTGGGTT 542
Qy 200 LeuAlaAlaCysPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySer 219
Db 543 CTTGGTGGCGCATGATGCTGTAGTCTCGGGGATATTTCTCGATCTGTAGTGGCTCC 602
Qy 220 PheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpVallys 239
Db 603 TATGATAATGAAGGAATTGTATCTTTTGCATGCTGCTTACTACTACATGTTGGATCAAG 662
Qy 240 SerVallysThrGlySerValPheThrTrpMetCysCysLeuSerTyrPheTyrMet 259
Db 663 GCAGTGAAGACTGGTTCCTATATGGGCTGCCAAGTGTGCCCTCGCTTATTTCTACATG 722
Qy 260 ValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheVal 279
Db 723 GTCTCTTCATGGGAGGCTATGTGTTCTTGTATCACTTATCTTCTACATGTTCTGCTG 782
Qy 280 LeuLeuLeuMetGlnAlaGlyTyrSerLysArgValTyrIleAlaTyrSerPheTyrIle 299
Db 783 CTAATGCTGACAGCCGCTTTTCTACCGGATCTAGTACCTACTGCTGCTTATTTTACTG 842
Qy 300 ValGlyLeuIleLeuSerMetGlnIleProPheValGlyPheGlnProIleArgThrSer 319
Db 843 CTGGGACCACTTCTTCTATGACAGATTCTTGTGTTTCCAGCCGCTCTTCTATCA 902
Qy 320 GluHisMetAlaAlaAlaGlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyr 339
Db 903 GAACACATGCCAGCTTTGGAGTGTGTCTCTCATCATCTGCTTTCTGATGATTAC 962
Qy 340 LeuArg-AspArgLeuThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLe 359

```

QY 713 LeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMetGlnLeuAspPhe 732
 Db 2018 CTAACGTCCTTATGACAAATGTTTACTACCGCTTTGGCAGGCTTACACAGAACGC 2077
 QY 733 ArgThrProGlyPheAspArgThrArgAsnAlaGluLeuGlyAsnLysAspLys 752
 Db 2078 AAGCGTCCACACAGGCTTTGACCGTGTTCGAATGCTGAGATTGTAATAGACTTTGAG 2137
 QY 753 PheLysHisLeuGluGluAlaPheThrSerGluHisTyrPheLeuValAlaGlyLeuVal 772
 Db 2138 CTTGATGCTCTGGAGAGCGTATACACAGACACTGGCTAGTCAGATATACAGGTA 2197
 QY 773 LysAlaProAspAsnArg 778
 Db 2198 AAGGACCTGGATAATCGA 2215

RESULT 11
 AY418286
 LOCUS
 DEFINITION Mus musculus ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY418286
 VERSION AY418286.1 GI:39774246
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2103)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 14671302
 2 (bases 1 to 2103)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 Location/Qualifiers
 1..2103
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>2103
 /gene="ITM1"
 /locus_tag="HCM6506"

gene
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.15e-198 Length: 2103
 Score: 2103.50 Matches: 405
 Percent Similarity: 74.00% Conservative: 96
 Best Local Similarity: 59.82% Mismatches: 142
 Query Match: 47.80% Indels: 35
 DB: 9 Gaps: 7

US-10-028-384-2 (1-826) x AY418286 (1-2103)

QY 105 TrpPheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsn 124
 Db 151 TACTTTAATATCGACTACCGGTTCTGCTCAGGAGGGGTTTATATAATCCATAAC 210
 QY 125 TrpPheAspGluAlaTrpTyrProLeuGlyArgLeuValGlyThrValTyrPro 144

Db 211 TGGTTTGTATGACCGGGCTTGGTACCCCTTGGCGCAATCATTTGAGGAGAACATTTACCCA 270
 QY 145 GlyLeuMetIleThrAlaGlyLeuIleHisTrpIleLeuAsnThrLeuAsnIleThrVal 164
 Db 271 GGTTTAATGATGATCACTTCTGTCGAATCTAGCAATCTTCCATTTCTTCCATATACATTT 330
 QY 165 HistLeuArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSer 184
 Db 331 GACATTCGGAATGCTGTGTTTCTGCGCCCACTTTCTCTCTTCCACACATCGTT 390
 QY 185 ThrPheLeuLeuThrArgGluLeuTrpAsnGlnGlyAlaGlyLeuLeuAlaCysPhe 204
 Db 391 AGTACCACTTACCAAGAGCTCAGGATGCGAGCTGGGCTTCTTGTGCTGCATG 450
 QY 205 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGly 224
 Db 451 ATTGCTGTAGTTCCTGGGTATATTTCTCGATCTGTAGCTGGCTCTTATGATATGAGGA 510
 QY 225 IleAlaIlePheAlaLeuGlnPheThrTyrTyrLeuTrpValLysSerValLysThrGly 244
 Db 511 ATTGCTATCTTTTGCATGCTGCTTACTTACTATGATGGATCAAGCAGTGAAGACTGGT 570
 QY 245 SerValPheTrpThrMetCysCysLeuSerTyrPheTyrMetValSerAlaTrpGly 264
 Db 571 TCCATCTATTTGGGCTGCCAGTGTGCCCTGGCTTATTTCTACATGGTCTCTTCATGGGA 630
 QY 265 GlyTyrValPheIleLeuAsnLeuIleProLeuHisValPheValLeuLeuMetGln 284
 Db 631 GGCTATGTGTTCTCGATCACTTGAATTCCTCATGCTCTGTGTGAATGCTGCACAGGC 690
 QY 285 ArgTyrSerLysArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuLeu 304
 Db 691 CGTTTCTCACCAGACTACGAGCTACTGTACTGTTTACTGCTGGCAGCACCATTCTT 750
 QY 305 SerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 324
 Db 751 TCTATGACAGATTTCTGTTGTTTCCAGCCGCTCTTTCATCAGAACACATGGCAGCC 810
 QY 325 AlaGlyValPheAlaLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeu 344
 Db 811 TTTGAGGTGTTGGTCTCTGTCAGATCCATGCTTTCGTAGATTACCTGCGCAGCATGTTG 870
 QY 345 ThrLysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaGlyAlaVal 364
 Db 871 AATCCACAGCAATTCGAAGTCTTTCCGGAGGTATTATCTCCCTG---GTTGGCTTTGTC 927
 QY 365 PheLeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTrpSerGlyArgPhe 384
 Db 928 CTCTCTACTGTGGAGCTCTCTCGA-AAA-----ATTCTCCCTGGACAGGCGCTTTC 980
 QY 385 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleLeuAlaSerValSer 404
 Db 981 TACTCTCTGCTGGATCCCTCTTATGCTAAGATAAATCCCATTTATGCTATTTCTTCT 1040
 QY 405 GluHisGlnProThrTrpValSerPhePheAspLeuHisIleLeuValCysThr 424
 Db 1041 GAGCACAGCCCAACACTGCTTCTTACTATTGATCTACAGCTCTCTGCTTCATG 1100
 QY 425 PheProAlaGlyLeuTrpPheCysIleLysAsnIleAsnAspGluArgValPheValAla 444
 Db 1101 TTTCCAGTTGGCTCTTATTACTGCTTTAGCAACGTGCTGCTGATGCTCGGATTTTATCATC 1160
 QY 445 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 464
 Db 1161 ATGTATGGTGGACCATGCTACTTTTCAGCTGTATGCTGCTGCTTATGCTGGTATTG 1220
 QY 465 ThrProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeu 484
 Db 1221 GCACCTGTATGTCAATTTCTGGCATTTGCTGCTTTCCAGGTGTGTGCCATATATG 1280
 QY 485 GlyAsp---AspMetLysArgGluAsnProProValGluAspSerSerAspGluAspAsp 503

QY 153 IleHisThrPheLeuAsnThrValHisIleArgAspValCysValPhe 172
 DB 474 ATCTACTGACATTCGGATTCCTCCGATTCCTTCATATCCCTGAGGTTTGTGTGA 533
 QY 173 LeuAlaProThrPheSerGlyLeuThrSerIleSerThrPheLeuLeuThrArgGluLeu 192
 DB 534 ACCGACCGCTTTTCGGCGTGAATACGACTCTCGTTCGGTATTCCTTTGGTAAAGAGCTT 593
 QY 193 TrpAsnGlnGlyAlaGlyLeuLeuAlaCysPheIleAlaIleValProGlyTyrIle 212
 DB 594 TGGGATACAGGAGCTGGTCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 653
 QY 213 SerArgSerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPhe 232
 DB 654 TCTCGTCTGTAGCTGGTTCGTATGATAACGAGCTGTGGATTTTCGCTCTGTGTCTG 713
 QY 233 ThrTyrTyrLeuTrpValLysSerValLysThrGlySerValPheThrMetCysCys 252
 DB 714 AGTTTATCTCTTTTAAAGCGGTGAACACTGGTTCATTAGCTTGGGCTCTTGGCTCT 773
 QY 253 CysLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeu 272
 DB 774 GGGTTTGGTACTTTTATATGTTTACGTTGGGAGGATATGTTTATATCACTTC 833
 QY 273 IleProLeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIle 292
 DB 834 GTCTCTCTTACGTTGGTGTGTGTATCATCCGGAAGGATTCGATCGAGGCTTTTACAT 893
 QY 293 AlaTyrSerThrPheTyrIleValGlyLeuIleLeuSerMetGlnIleProPheValGly 312
 DB 894 GCTTATACGTATGATATCTTGGGATGTGTCTCGGATGCAATTCGCTTGTGGTT 953
 QY 313 PheGlnProIleArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuGln 332
 DB 954 TTC---ACATGTTTCATCGGTGAACATATGGGTGCCATGGGTCTCTTCTTATTATGTCAG 1010
 QY 333 AlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGlu---PheGlnThr 351
 DB 1011 GTGTTTACTTCTGGACGTGGTAACTACAGCTCAATGACCAAGTGTTCCTCAACG 1070
 QY 352 LeuPheLeuGlyValSerLeuAla-----AlaGlyAlaValPheLeuSerValIle 369
 DB 1071 TTTTTCGAGAATAACTGTGACATCGCTATCTCTGTTGGTGTGTGTGTGTGTGTGTGT 1127
 QY 370 TyrLeuTyrThrGlyTyrIleAlaProThrSerGlyArgPheTyrSerLeuTrpAsp 389
 DB 1128 ---GGACAGCGTCTGGCTATATCTCATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1184
 QY 390 ThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThr 409
 DB 1185 CGACTTACGCAAGGATCATATTCCTATTATTCATCTGTCTGTGACATCAGCCCTACT 1244
 QY 410 ThrTrpValSerPhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeu 429
 DB 1245 GCTGGTCTGATTCATGTTTACCTACCATATCTGCTCTCTCTCTCTCTCTCTCTCTCT 1304
 QY 430 TrpPheCysIleLysAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSer 449
 DB 1305 TACTTCTGTTTCAACGCTTTCGCGATCTACCATATTTATTTGTCATGTATGTCTCACT 1364
 QY 450 AlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCys 469
 DB 1365 ACCTTGTACTTGT 1424
 QY 470 MetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGlyAspMetLys 489
 DB 1425 CTATCAGTGCCTACTGTCTCTGCCACTATCAAGATTATTA-----ACTTCTCTGTGA 1478
 QY 490 ArgGluAsnProProValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsn 509
 DB 1479 AGCAACAAACAAAGGTTCTCAGACTGTCTC-----ACGAAGAGAGCT 1523
 QY 510 LeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGluLysThrGluGlu 529

DB 1524 GGTAGTTCAAAGCTTCTTCA-----AAGGTTACACTTGTAGCTCTCAGCCTTCCAG 1577
 QY 530 GlyLeuGlyProAsnIleLysSerIleValThrMetLeu---MetLeuMetLeuLeuMet 548
 DB 1578 AAGAATGCT-----GCCATTGCTCTTCTTGTGTGTGTATTTATTGCTCAGT 1625
 QY 549 MetPheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerValVal 568
 DB 1626 AGATATGATCTACTCATCGATGGGTACAGAGAGCATATTCATCTCCCTCAATGTC 1685
 QY 569 LeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAspAspPheArgGluAlaTyr 598
 DB 1686 TTAGCTCAAGAGGAGCCCGGACAGAAATCATCTTTGATGATTACCGGAGGCATAC 1745
 QY 599 PheTrpLeuArgGlnAsnThrAspGluHisAlaArgValMetSerTrpTrpAspTyrGly 608
 DB 1746 TACTGGCTTAGGCAAAACACTGCCACTGTAGCTAAGATTATGTCTGGTGGGACTATGG 1805
 QY 609 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 628
 DB 1806 TACCAATCACTGCCATGGGAATAGAACTGTCTCATCTGATTAACATACCTCGAACAAC 1865
 QY 629 SerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGlnThrAlaTyrLysIle 648
 DB 1866 ACTCACATTCGTACCGTTGGACGCCCATCTCTTTATGAAGATGATGCGTATGACATC 1925
 QY 649 MetArgThrLeuAspValAspTyrValLeuValIlePheGlyValIleGlyTyrSer 668
 DB 1926 ATGAGTTCACCTGATGTGAATATGTTGCTCTTGTGTGTGTGTGTGTGTGTGTGT 1985
 QY 669 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 688
 DB 1986 TCAGATGATATCAACAAGTTCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2045
 QY 689 AspIleArgGluSerAspTyrPheThrProGlnGlyGluPheArgValAspLysAlaGly 708
 DB 2046 ---ATCAAGAACCTGATTACCTTTGTG---AATGGTGAATTCGTTAGACAAAGTGCA 2099
 QY 709 SerProThrLeuLeuAsnCysLeuMetTyrLysMetSerTyrTyrArgPheGlyGluMet 728
 DB 2100 TCACCGAAGTGTGAACCTGTCTCATGATCAAGTTATGCTATTACAGGTTCCGTGTAGCT 2159
 QY 729 GlnLeuAspPheArgThrProProGlyPheAspArgThrArgAsnAlaGluIleGlyAsn 748
 DB 2160 ACCAGAAATACGCAAGCCCGGATGATGATAGAGAGAGAGAGAGATAGGAAAC 2219
 QY 749 LysAspIleLysPheLysHisLeuGluGluAlaPheThrSerGluHisTrpLeuValArg 768
 DB 2220 AAGATATCAAACTTGAACTTAGAAGAGCTTAGACACATCCACATCCCACTGGATTCGT 2279
 QY 769 IleTyrLysValLysAlaProAspAsnArg 778
 DB 2280 ATTTACAGAGTTAAACCTCCTACAAATAGG 2309
 RESULT 13
 AY418285 2079 bp DNA linear GSS 17-DEC-2003
 LOCUS Pan troglodytes ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 AY418285
 VERSION AY418285.1 GI:39774245
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE 1 (bases 1 to 2079)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITILE Inferring nonneutral evolution from human-chimp-mouse orthologous

gene trios
 Science 302 (5652), 1960-1963 (2003)
 14671302
 2. (bases 1 to 2079)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kefauver, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cartagalli, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1. 2079
 Location/Qualifiers
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..2079
 /gene="ITIM1"
 /locus_tag="HMC6506"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,04e-157 Length: 2079
 Score: 1691.00 Matches: 333
 Percent Similarity: 61.19% Conservative: 77
 Best Local Similarity: 49.70% Mismatches: 228
 Query Match: 38.42% Indels: 32
 DB: 9 Gaps: 3
 US-10-028-384-2 (1-826) x AY418285 (1-2079)
 QY 106 PheAsnTyrArgSerThrHisLeuAlaSerHisGlyPheTyrGluPheLeuAsnTrp 125
 DB 154 TTTAAATATCGACTACGANNCCCTGGCTGNNNGGGTTTATAAATTCATACTGG 213
 QY 126 PheAspGluArgAlaTyrProLeuGlyArgGlyValGlyGlyThrValTyrProGly 145
 DB 214 TTTGATGACCGAGCTGGTACCTTTGGGACGAATCATTTGGAGGANNNNTTACCAGGT 273
 QY 146 LeuMetIleThrAlaGlyLeuIleHisTyrPheLeuAsnThrLeuAsnIleThrValHis 165
 DB 274 TTAATGATACCTTCTGCGAATACCAATGACTCCATTTTCCACATCACCATCGAC 333
 QY 166 IleArgAspValCysValPheLeuAlaProThrPheSerGlyLeuThrSerIleSerThr 185
 DB 334 ATTCGGAATGCTNTGTGTTCCTGGGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCAG 393
 QY 186 PheLeuLeuThrArgGluLeuTyrAsnGinglyAlaGlyLeuLeuAlaAlaCysPheIle 205
 DB 394 TACCACCTTACCAAGAGCTCAAGATGAGGAGGCTGGGCTTCTTCTGCTGCTGCTGAT 453
 QY 206 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerPheAspAsnGluGlyIle 225
 DB 454 GCTGTAGTCTCTGGATATATCTCCGATCTGTGGCTGCTCTCTATATATCAAGAGGAT 513
 QY 226 AlaIlePheAlaLeuGlnPheThrTyrTyrLeuTyrValIleSerValIleThrCysSer 245
 DB 514 GCCATCTTTGCGAGTACTACCTACTACTGATGATGATGATGATGATGATGATGATGAT 573
 QY 246 ValPheTrpThrMetCysCysCysLeuSerTyrPheTyrMetValSerAlaTrpGlyGly 265
 DB 574 ATCTGTGGGAGCAGTAAAGTGTGCTTGTCTTATTTCTATCTGCTGCTGCTGCTGCTG 633
 QY 266 TyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuMetGlnArg 285
 DB 634 TATGTGTTCTGATCACTTAATTTCTCTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 693
 QY 286 TyrSerIleArgValTyrIleAlaTyrSerThrPheTyrIleValGlyLeuIleLeuSer 305
 DB 694 TTCTCTACCGGAGTCTATGTGGCTACTGTACTGTCTTACTGCTGCTGCTGCTGCTGCTTCT 753

QY 306 MetGlnIlePropheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaA 325
 DB 754 ATGCAGATCTCTTTGGGTTTCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 813
 QY 326 GlyValPheAlaLeuLeuGlnAlaTyrAlaPheLeuGlnTyrLeuArgAspArgLeuThr 345
 DB 814 NNN 873
 QY 346 LysGlnGluPheGlnThrLeuPhePheLeuGlyValSerLeuAlaAlaGlyAlaValPhe 365
 DB 874 CCANACAAATTTNNNGTTCTTTTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 933
 QY 366 LeuSerValIleTyrLeuThrTyrThrGlyTyrIleAlaProTyrPheSerGlyArgPheTyr 385
 DB 934 NNN 993
 QY 386 SerLeuTyrAspThrGlyTyrAlaIleHisIlePheIleAlaSerValSerGlu 405
 DB 994 TCACTGCTGGATCTCTTATGCTTAAGAACAAATCCCATCATCTGTTCTGTGCTGAG 1053
 QY 406 HisGlnProThrTyrTrpValSerPhePheAspLeuHisIleLeuValCysThrPhe 425
 DB 1054 CATCAGCCCAACCTGCTCTCATCTATTTCACCTNNAGCTCTCTGCTCTCATGTTT 1113
 QY 426 ProAlaGlyLeuTyrPheCysIleLeuAsnIleLeuAspGluArgValPheValAlaLeu 445
 DB 1114 CAGTTGGCTCTATTACTGTTNGCACTGCTGATGCCCGGATTTTATCATCATG 1173
 QY 446 TyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThr 465
 DB 1174 TATGCTGTGACCAACCATGCTACTTTTCAGCTGTAATGNNNNNNNNNNNNNNNNNN 1233
 QY 466 ProValValCysMetLeuSerAlaIleAlaPheSerAsnValPheGluHisTyrLeuGly 485
 DB 1234 NNN 1290
 QY 486 AspAspMetLysArgGluAsnProProValGluAspSerSerAspGluAspLysArg 505
 DB 1290 ----- 1290
 QY 506 AsnGlnGlyAsnLeuTyrAspLysAlaGlyLysValArgLysHisAlaThrGluGlnGlu 525
 DB 1291 -----AAGAAATGACATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1329
 QY 526 LysThrGluGluGlyLeuGlyPro-----AsnIleLysSerIleValThrMetLeu 542
 DB 1330 AAGNNNNAGGATTCGCCNCCCTATTAAAGAAATGAAGTGGCAAGTGGGATGATGATGCTC 1389
 QY 543 MetLeuMetLeuMetMetPheAlaValHisCysThrTyrValThrValThrSerAsnAlaTyr 562
 DB 1390 ATGGCTTTCTTCTCATCTACCTACCTTTCATTTCACTGGGTGACCATGAGGCTCTAC 1449
 QY 563 SerSerProSerValValLeuAlaSerTyrAsnHisAspGlyThrArgAsnIleLeuAsp 582
 DB 1450 TCTTCTCGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1509
 QY 583 AspPheArgGluAlaTyrPheThrLeuArgGlnAsnThrAspGluHisAlaArgValMet 602
 DB 1510 GACTTCCGAGAGCATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1569
 QY 603 SerTrpTyrPheTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAsp 622
 DB 1570 NCTGCTGGGATTTATGGCTATCAGATACAGTATGCGCAACCCGACAAATTTTAGTGGAC 1629
 QY 623 AsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerAsnGlu 642
 DB 1630 AATAACATGGAATATATACCATATTTCTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1689
 QY 643 ThrAlaAlaTyrLysIleMetArgThrLeuAspValAspTyrValLeuValIlePheGly 662
 DB 1690 GAAAAGGCTATGATCATGAGGAGCTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTG 1749
 QY 663 GlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArgIle--- 681

```

1750 GGCCTCACCTGGGTATCTCTCATGATATCAACAGTTCNNNGATGGTCCGGATTGGA 1809
QY 682 AlagluGluHisProLysAspIleArgGluSerAspTyrPheThrProGlnGlyGlu 701
Db 1810 GGGAGNACAGANNANGCAACATATCAAGGAGATGACTATTATATATCCACTGGGGAG 1869
QY 702 PheArgValAspIleGlySerProThrLeuLeuAsnCysLeuMetTyrLysMetSer 721
Db 1870 TTCGGTGTGNNCNTGAAGGTCTCCAGTGTGCTCCANNNNNNNNNNNNNNNNNNNN 1929
QY 722 TyrTyrArgPheGlyGluMetGlnLeuAspPheArgThrProGlyPheAspArgThr 741
Db 1930 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1989
QY 742 ArgAsnAlaGluIleGlyAsnLysAspIleLysPheLysHisLeuGluGluAlaPheThr 761
Db 1990 CGAAATGCTGAGATGGGAATAAGACTTTCAGCTTGATGCTCGGAGGAGCATATACC 2049
QY 762 SerGluHisThrLeuValArgIleTyrLys 771
Db 2050 ACAGAACATTCGCTGGTCAGGATATACAG 2079

```

RESULT 14

```

BUS15954
LOCUS
DEFINITION
AGENCOURT 10120433 NIH MGC 134 Mus musculus cDNA clone
IMAGE:6512642 5', mRNA sequence.
BUS15954
VERSION
BUS15954.1 GI:22823480

```

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

Mus musculus
Mus musculus (house mouse)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 950)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaapbs@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14085 row: b column: 03
High quality sequence stop: 680.
Location/Qualifiers
1..950
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6512642"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 134"
/notes="Vector: pCMV-Sport6.1; Site 1: EcoRV; Site 2: NotI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
this is a NIH_MGC Library."

```

FEATURES

source

```

Alignment Scores:
Pred. No.: 3,71e-145 Length: 950
Score: 1563.00 Matches: 300
Percent Similarity: 96.23% Conservative: 6
Best Local Similarity: 94.34% Mismatches: 10
Query Match: 25.51% Indels: 2
DB: 5 Gaps: 0

```

ORIGIN

```

RESULT 15
CR610915
LOCUS
DEFINITION
full-length cDNA clone CS0DK010F001 of HeLa cells Cot 25-normalized
of Homo sapiens (human).
ACCESSION
CR610915
VERSION
CR610915.1 GI:50491722

```

US-10-028-384-2 (1-826) x BUS15954 (1-950)

```

QY 215 SerValAlaGlySerPheAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 234
Db 1 TCAGTGGCGGGATCTTTGATATGAAGGATGGCCATTTTGGCTTCAGTTTCATTAC 60
QY 235 TyrLeuTrpValLysSerValLysThrGlySerValPheTrpThrMetCysCysLeu 254
Db 61 TACTTATGGTAAAGTCTGTGAAGACCGGGTCTGTCTTCTGGCAATATGCTGCTGCTTG 120
QY 255 SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleLeuLeuIlePro 274
Db 121 TCATATTCTACATGGTCTCTGGTGGGAGGTATGTGTTCATCATCACTCATCCCT 180
QY 275 LeuHisValPheValLeuLeuLeuMetGlnArgTyrSerLysArgValTyrIleAlaTyr 294
Db 181 CTCCTATGTGTTGTGTTGCTGCTGATGAGAGGTACAGCAAGAGAGTCTACATAGCATAT 240
QY 295 SerThrPheTyrIleValGlyLeuLeuSerMetGlnIleProPheValGlyPheGln 314
Db 241 AGCATTCTACATTTGGGTTTAATATTATCCATGAGATACCTTTTGTGGATTTTCAG 300
QY 315 ProfileArgThrSerGluHisMetAlaAlaGlyValPheAlaLeuLeuGlnAlaTyr 334
Db 301 CCAATCAGAAACAAGCAGCACATGGCAGCTGCAGGTGCTCTTTCGGCTGCTGCAAGCTTAC 360
QY 335 Ala-PheLeuGlnTyrLeuArgAspArgLeuThrLysGlnGluPheGlnThrLeuPhePh 354
Db 361 GCTTTTTCGAGTATCTGAGAGACCGGTGACAAACAGGAGTTCAGACCCCTTTTCTT 420
QY 354 eLeuGlyValSerLeuAlaAlaGlyAlaValPheLeuSerValIleTyrIleuThrTyrTh 374
Db 421 TTTGGGTGCTCTCCTAGCTCAGCGCTGTGTTCTTAGTGTCTATCTATCTGACATACAC 480
QY 374 rGlyTyrIleAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaL 394
Db 481 AGTTATATTGCACCATGGAGTGGCAGGTTTATTCTACTATGGGATCTGGGTATGCATA 540
QY 394 sIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpValSerPh 414
Db 541 AATACACATTCCAATTATTGTCATCAGTGTCTGAACATCAGCCTPACGACATGGGTGCTTT 600
QY 414 ePhePheAspLeuHisIleLeuValCysThrPheProAlaGlyLeuTrpPheCysIleL 434
Db 601 CTTCTTGTGATCATATTTCTGTATGTACTCTTCCAGCAGGCTATGGTCTGCAATCAA 560
QY 434 sAsnIleAsnAspGluArgValPheValAlaLeuTyrAlaIleSerAlaValTyrPheAl 454
Db 661 AAATATCAACGATGAAAGAGTGTGTCGCTCTGTATGCGATCAGTGTGCTTACTTTGC 720
QY 454 aGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuSerAlaI 474
Db 721 CNGAGTGTATGGTGGGCTGATGTGACTCTGACCCCGGTCTGTGATGCTGTGGGGCT 780
QY 474 eAlaPheSerAsnValPheGluHisTyrLeuGlyAspAspMetLysArgGluAsnProPr 494
Db 781 CGCTCTCTCCAATGTTTTTTGACACTATTTGGGGGATGACATTTGAAGGAAAAAACCACC 840
QY 494 oValGluAspSerSerAspGluAspLysArgAsnGlnGlyAsnLeuTyrAspLysAl 514
Db 841 TGTGGAGGACGACACTGATGAGATGACAAAGAAACCCCGGAAACTTTGTATGACAAAGCC 900
QY 514 aGlyLysValArgLysHisAlaThrGluGln-GluLysThrGluGluGly 530
Db 901 AGGTAAGTGAAGAAGCATGTGACCAGACAAAGAGAAACCTGAACAGGA 950

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 7954.53 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 718

Sequence: 1 MGSDSCVLSVFQTLKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues.

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131645_6800/app_query.fasta_1.3740
-DB=GenEmbl -QFMT=fastap -SUFFIX=Oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384 @CEN 1 1 20362 @runat_14122004_131645_6800 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 718 | 100.0 | 2733 | 6 | AX799086 Sequence |
| 2 | 718 | 100.0 | 2733 | 8 | D28952 Saccharomyc |
| 3 | 631 | 87.9 | 3633 | 8 | YSCSTT3 |
| 4 | 605 | 84.3 | 2000 | 6 | AX595714 Sequence |

| | | | | | | | |
|------------|------------|------------|------------------------------------|---------|-----|--------|-----------------|
| RESULT 1 | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| LOCUS | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| DEFINITION | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| ACCESSION | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| VERSION | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| KEYWORDS | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| SOURCE | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| ORGANISM | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| REFERENCE | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| AUTHORS | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| TITLE | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| JOURNAL | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| FEATURES | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| source | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| ORIGIN | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |

Alignment Scores:
Pred. No.: 2733
Score: 718.00 Matches: 718

AX819898 Sequence
AX830928 Sequence
C0450623 Sequence
X87672 S.cerevisia
Continuation (11 o
AE016903 Eremothec
AX489023 Sequence
AB015232 Schizosac
AB034353 S.pombe c
Continuation (4 of
CR30947 Candida g
Continuation (22 o
AF132552 Drosophil
AX799088 Sequence
C0589353 Sequence
C0596819 Sequence
C0589352 Sequence
AC018145 Drosophil
AC007853 Drosophil
AC008206 Drosophil
AB003750 Drosophil
BX842635 Neurospor
AF548377 Talaromyc
AX079689 Sequence
CO782554 Sequence
BD127193 Primer fo
AK074587 Homo sapi
BC003206 Mus muscu
AX799082 Sequence
AY074880 Homo sapi
AX799084 Sequence
C0573675 Sequence
BC052433 Mus muscu
C0573674 Sequence
U13019 Caenorhabdi
AC104631 Drosophil
AC104643 Homo sapi
BX255961 Zebrafish
AC092024 Homo sapi

ALIGNMENTS

| | | | | | | | |
|------------|------------|------------|------------------------------------|---------|-----|--------|-----------------|
| RESULT 1 | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| LOCUS | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| DEFINITION | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| ACCESSION | AX799086 | AX799086 | Sequence 5 from Patent WO03054008. | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| VERSION | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| KEYWORDS | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| SOURCE | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| ORGANISM | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| REFERENCE | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| AUTHORS | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| TITLE | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| JOURNAL | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| FEATURES | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| source | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |
| ORIGIN | AX799086.1 | AX799086.1 | GI:37605060 | 2733 bp | DNA | linear | PAT 08-OCT-2003 |

Alignment Scores:
Pred. No.: 2733
Score: 718.00 Matches: 718

| | | | | | | | |
|---|------|--|------|-----------------|--|--|--|
| Percent Similarity: 100.00% | | | | Conservative: 0 | | | |
| Best Local Similarity: 100.00% | | | | Mismatches: 0 | | | |
| Query Match: 100.00% | | | | Indels: 0 | | | |
| DB: 6 | | | | Caps: 0 | | | |
| US-10-028-384-6 (1-718) x AX799086 (1-2733) | | | | | | | |
| Qy | 1 | MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle | 20 | | | | |
| Db | 219 | ATGGGATCCGACCGGCTGGTGTGTTTCTGTCTGTTTCAGACCACTCCTCAAGCTGTCTATC | 278 | | | | |
| Qy | 21 | PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu | 40 | | | | |
| Db | 279 | TTCGTGGCGATTTTGGGCTGCCATATCATACGTTTGTTCAGCATCAAAATTGAG | 338 | | | | |
| Qy | 41 | SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysTyrLeuValAsn | 60 | | | | |
| Db | 339 | TCTATTATCCATGAATTCGACCCCTGGTTCATATATAGGCTACCAAAATATCTGTCAAC | 398 | | | | |
| Qy | 61 | AsnSerPheTyrLysPheLeuAsnTTPheAspAspArgThrTTPyrProLeuGlyArg | 80 | | | | |
| Db | 399 | AATTCGTTTACAGTTTGTGAATCTGTTTGACGACCGTACCTGGTACCCCTCGGAAGG | 458 | | | | |
| Qy | 81 | ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla | 100 | | | | |
| Db | 459 | GTTACTGGAGGACCTTTATATCTCTGTTTGTATGACGACTAGTGGTTCATCTGGCACGCC | 518 | | | | |
| Qy | 101 | LeuArgAsnTTPheLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro | 120 | | | | |
| Db | 519 | CTGGGCAACTGGTGGGCTGGCCATTTGACATCAGAAACGTTTGTGTGTATTTGGGCCA | 578 | | | | |
| Qy | 121 | LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla | 140 | | | | |
| Db | 579 | CTATTTCCTGGGGTCAACCGCTGGGCACTTACGAAATTTACGAAAGAGATTAAGATGCC | 638 | | | | |
| Qy | 141 | SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer | 160 | | | | |
| Db | 639 | AGCGCTGGGCTTTTGGCTGTGTTTATAGCCATTTGTCCCGTTATATATCTAGATCA | 698 | | | | |
| Qy | 161 | ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet | 180 | | | | |
| Db | 699 | GTGGCGGGGCTCTACGATAATAGGCCATTTGCCATTTACATATAATGGTCTCTTTCTATG | 758 | | | | |
| Qy | 181 | PheTTPileLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe | 200 | | | | |
| Db | 759 | TTTTGGATTAGGCCCAAGACCTGGCTCTATCATGCACGCAACGTTGCAGCTTTATTC | 818 | | | | |
| Qy | 201 | TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuIleProLeu | 220 | | | | |
| Db | 819 | TACTTCTACATGGTGTGGGCTGGGGTGGGATCGTGTTCATACCAACCTGTATCCCACTC | 878 | | | | |
| Qy | 221 | HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr | 240 | | | | |
| Db | 879 | CATGCTCTTTTGTGTGATTTTGATGGGCAGATATTCGTCGCAACTGTATTCGCTACACC | 938 | | | | |
| Qy | 241 | ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro | 260 | | | | |
| Db | 939 | ACTTGGTACGCTATTGGAACCTTTGTCATCCATCCAGATCCCATTTCTCGGTTTCTACCT | 998 | | | | |
| Qy | 261 | IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla | 280 | | | | |
| Db | 999 | ATCAGGTCTACAGCACCATGCGCCGCTTGGGTGTTTTCGGTTTGTATTCAGATTGTGCGC | 1058 | | | | |
| Qy | 281 | PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal | 300 | | | | |
| Db | 1059 | TTCGGTGACTTTCGTGAAGGGCCAAATCAGCACAGCTAAAGTTTAAAGTCACTATGATGTT | 1118 | | | | |
| Qy | 301 | SerLeuPheLeuIleLeuValLeuGlyValAlaGlyLeuSerAlaLeuThrTyrMetGly | 320 | | | | |
| Db | 1119 | TCTCTGTTTGTGATCTTGGTCCCTGGTGGTGGGACTTCTCGCTTGTACCTATATGGG | 1178 | | | | |
| Qy | 321 | LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle | 340 | | | | |

| | | | |
|----|------|---|------|
| Db | 1179 | TTGATTGCCCTTGGACTGGTAGATTATTTATCGTTATGGGATACCAACTACGCAAGATC | 1238 |
| Qy | 341 | HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe | 360 |
| Db | 1239 | CACATTCCTATCATTTGGCTTCGTTCCGACATCAACCCGTTTCGTGGCCGCTTTCTTC | 1298 |
| Qy | 361 | PheAspThrHisPheIleThrLeuPheProAlaGlyValPheLeuLeuPheLeuAsp | 380 |
| Db | 1299 | TTTGATACCACTTTTGTGCTGCTATTCGCCGCCGCTGTATTCCTACTATTCTTCGAC | 1358 |
| Qy | 381 | LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly | 400 |
| Db | 1359 | TTGARAAGACGACGACGTTTGTGTCATCCGTTACTCCGTTCTGTCTGCTACTTTGCCGT | 1418 |
| Qy | 401 | ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla | 420 |
| Db | 1419 | GTTATGTTTAGATTGATGTTGACCTTTGACACGACGTCATCTGTGTCCGCCGCGCTCGCA | 1478 |
| Qy | 421 | LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle | 440 |
| Db | 1479 | TTGTCCAGATATTTGACATCTACCTGGATTTCAAGACAGTGAACGCAAAATAGCCATC | 1538 |
| Qy | 441 | LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr | 460 |
| Db | 1539 | AAACCTGCCGCACTACTGGCCAAATTTGATTTTCCGGATCAATCACTTTTATTTGAT | 1598 |
| Qy | 461 | LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal | 480 |
| Db | 1599 | CTTTTCGCTCTTCCATTTCTACTTGGGTAAACAAGACTGCATACTCTCTCTCTCTGTTT | 1658 |
| Qy | 481 | LeuProSerGlnThrProAspGlyLysLeuAlaIleLeuLeuAspPheAspGluAlaTyr | 500 |
| Db | 1659 | TTGCCATCAACAACCCAGATGGTAAATTTGGCGTGTATCGACGACTTTCAGGAAGCGTAC | 1718 |
| Qy | 501 | TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaIleTTPAspTyrGly | 520 |
| Db | 1719 | TATTTGGTTAAGAAATGAATCTGATGAGGACAGTAAGGTTGCAGCGTGTGGATTACGGT | 1778 |
| Qy | 521 | TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTTPAsnAsn | 540 |
| Db | 1779 | TACCAAAATTTGGTGGATGGCAGACAGAACCACTTTAGTCGATATAACAACACGTTGAACAAT | 1838 |
| Qy | 541 | ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle | 560 |
| Db | 1839 | ACTCACATCGCCATCGTTGGTAAAGCCATGCTTCCCTCCCTGAAGAGAAATCTTACGAAAT | 1898 |
| Qy | 561 | LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly | 580 |
| Db | 1899 | CTAAAAGACATGATGTCGATTATCTTGGTTCATCTTTTGGTGTCTTAATTTGGGTTTGGT | 1958 |
| Qy | 581 | GlyAspAspIleAsnLysPheLeuTTPMetIleArgIleSerGluGlyIleTTPProGlu | 600 |
| Db | 1959 | GGTGTATGACATCAACAATTTCTTGGTATGATCAGAAATTCAGCGGGAATCTGCGCCAGAA | 2018 |
| Qy | 601 | GluIleLysGlyLysArgTyrPheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAla | 620 |
| Db | 2019 | GAGATAAAAGACGCTTATTTCTATACCGCAGGAGAGAAATACAGAGTAGATGCAAGGCT | 2078 |
| Qy | 621 | SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu | 640 |
| Db | 2079 | TCTGACACCATGAGGAATCTCGCTACTTTTACAAGATGCTCTACAAGATTTTCCCAACAATTA | 2138 |
| Qy | 641 | PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal | 660 |
| Db | 2139 | TTCAATTTGGTGGCCAGCCACTGACAGAGTGGCTCAACAAATGATCACCATTAGACGTC | 2198 |
| Qy | 661 | ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTTPMetValArgIleTyr | 680 |
| Db | 2199 | CCACCATTAGACTACTTTCGACGAGCTTTTACTTCCGAAAACTGGATGGTTAGAAATATAT | 2258 |
| Qy | 681 | GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg | 700 |
| Db | 2259 | CAATTGAAGAGAGATGATGCCCAAGGTAGACTTTTGAGGGAGCGTTGTGAGTTAACCCAGG | 2318 |

Qy 701 SerSerThrIysThrArgSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2319 TCTTCTACGAAACCAAGAGGTCCTAAAGAGACCTGAATAGGCTTGAGATC 2372

RESULT 2
 YSCSTT3
 LOCUS YSCSTT3 2733 bp DNA linear PLN 18-JAN-2003
 DEFINITION Saccharomyces cerevisiae gene for STT3 protein, complete cds.
 ACCESSION D28952
 VERSION D28952.1 GI:468493
 KEYWORDS
 SOURCE
 ORGANISM
 Saccharomyces cerevisiae (baker's yeast)
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE
 1 Yoshida, S., Ohya, Y., Nakano, A. and Anraku, Y.
 STT3, a novel essential gene related to the PKC1/STT1 protein
 Kinase pathway, is involved in protein glycosylation in yeast
 Gene 164 (1), 167-172 (1995)
 96060859
 7590309
 2 (bases 1 to 2733)
 YSCSTT3
 DIRECT SUBMISSION
 YOSHIDA, S.
 TITLE
 Submitted (14-MAR-1994) Satoshi Yoshida, Faculty of Science,
 University of Tokyo, Department of Biology, 7-3-1, Hongo,
 Bunkyo-ku, Tokyo 113, Japan (Tel:81-3-3812-2111 (ex.4461),
 Fax:81-3-3812-4929)
 LOCATION/QUALIFIERS
 1. .2733
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /strain="S288C"
 /db_xref="taxon:4932"
 /clone_lib="pSTS3-105"
 /library="Library of T. Yoshida and Y. Anraku"
 219. .2375
 /genes="STT3"
 219. .2375
 /genes="STT3"
 /codon_start=1
 /product="STT3 protein"
 /protein_id="BA08079.1"
 /db_xref="GI:603808"
 /translation="WGSRRSCVLSVFOTILKLVFAFGAAISSELFVAVIKPESIIH
 EFDWPNRYATKLVNNGFYKFLNWFDDRTWYPLRGVTGTLPLGLTTSATFTHALR
 NMLGFLDIRNVCVLPFSGVTAWATFEFTKIDASAGLLAAGFIALVPGVIGRS
 VAGSYDNEAIAITLLMTVFMFWIKAKQKTSIMHATCAALFYVMWSWAGGVFTNL
 PLHVFLLILMGYSKLYSAYTTVAICTVASMOTPFVGFPLIRSNHMAALGVFGLI
 QIVAGDFVKGQISTAKSKVIMWSLFIILVGVGLSALTVMGLIAPMTGRFYSGLD
 TVYAKHIFILIASVSEHQVSWPAFFEDTHFLWLPFAGVFLFLDLKDHVFIAYS
 VICSIFAGVNRMLTLPVICSAAVALSKIFDIYDFKTSRDKYAIKPAALIALKI
 VSGSFIFLYLRFVHTSWTRTAYSSPVLPSQTPDGKALIDDREAYVLRMSD
 ESKVAWMDYQYQGMADRTLLVDNNTWNTHTAIVGKAMASPEKSEYILKEHDV
 DYVLVFLGIGGEDINKFLMIRISEGIVPEEIKERYFYTAEGEYRVDARASETM
 RNSLYKNYSKDFPOLFNFGQATDRVQOMITPLDVPFLDYFDEFTSENWVRIVQL
 KKDDAQCRILRDVGLSTRSSTKTRSIKRPGLGRV"

FEATURES

source

gene

CDS

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2733
 Score: 718.00 Matches: 718
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x YSCSTT3 (1-2733)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValle 20
 Db 1299 TTTGATACCCCACTTTTGTATTCCTCCCGCGGTGTATTCCTACTATTCCTGCAC 1358

Db 219 ATGGATCCGACCGTCGTGTGTTTGTCTGTGTTTTCAGACCATCTCAGCTCGTCATC 278
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 279 TTCGTGGCGATTTTGGGGCTGCCATATCATCATCGTTTGTTCAGCTCATCAAAATTGAG 338
 Qy 41 SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysTyrIleuValAsn 60
 Db 339 TCTATTATCCATGATTCGACCCCTGGTCAATATAGGGCTACCAAAATATCTCGTCAAC 398
 Qy 61 AsnSerPheTyrLysPheLeuAsnTTPheAspAspArgThrTTPyrProLeuGlyArg 80
 Db 399 AATTCGTTTTTACAAGTTTTTGAACCTGTTTGGACGACCTACCTGGTACCCCTCGAAGG 458
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
 Db 459 GTTACTGGAGGACTTTATATCTGTTGATGACGACTAGTGGTTCATCTGGCAGCGC 518
 Qy 101 LeuArgAsnTTPLeuGlyLeuProlIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 519 CTGGCGCAACTGGTTGGGCTTGGCCATTCACATCAGAAACGTTTGTGTATTTGGCGCA 578
 Qy 121 LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 Db 579 CTATTTTCTGGGGTCACCGCTGGGGACTTACGAAATTTACGAAAGAGATTAAGATGCC 638
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 639 AGCGCTGGGCTTTTGGCTGTGTTTATAGCCATTGCCCCGGTTATATATCTAGATCA 698
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 699 GTGGCGGGCTCTACGATATATAGGCCATTGCCATTACACTATTAATGTCTCACTTTCATG 758
 Qy 181 PheTTPleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 759 TTTTGGATTAAAGGCCAAAGACTGGCTCTATCATGACGACAGTGTGCAGCTTTATTC 818
 Qy 201 TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 819 TACTTCTACATGGTGTGGCTTGGGGTGGATACGTGTTTCATACCAACTTGTATCCACTC 878
 Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysIleuTyrSerAlaTyrThr 240
 Db 879 CATGTCCTTTTGTGATTTGATGGGCGAGATATCTGCAAACTGATTTCTGCCATACACC 938
 Qy 241 ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 939 ACTTGGTACGCTATTGAACTGTTCATCCATCCAGATCCCATTTGTCGGTTTCTCTACT 998
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 999 ATCAGTCTAACGACCACTTGGCGCATTTGGGTGTTTTCGGTTTGAATTCAGATTGTCCGC 1058
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 1059 TTCGGTCACTTCGTGAAGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGGTT 1118
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 1119 TCTCTGTTTTTGATCTTGGTCTTGGTGGTGGACTTTCTGCCCTTACCTATATGGGG 1178
 Qy 321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle 340
 Db 1179 TTGATTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACCAAGATC 1238
 Qy 341 HistleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe 360
 Db 1239 CACATTCCTATCATTTGCTCGTTTCGGAACATCAACCGTTTCGTGGCGCGCTTCTTC 1298
 Qy 361 PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1299 TTTGATACCCCACTTTTGTATTCCTCCCGCGGTGTATTCCTACTATTCCTGCAC 1358

Db 1253 ATGGATCCGACCGGTCGTGTTTGTCTGTGTTTTCAGACCATCTTCAAGTCGTATC 1312
Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
Db 1313 TTGCGTGGCGATTTTGGGGCTGCCATATCATCAGCTTTGTTGCGATCATCAATTTGAG 1372
Qy 41 SerIleIleHisGluPheAspProTrrPheAsnTyrArgAlaThrIleLysTyrLeuValAsn 60
Db 1373 TCTATTATCCATCAATTCGACCCCTGGTTCAATTATAGGGCTACCAATATCTCGTCAAC 1432
Qy 61 AsnSerPheTyrLysPheLeuAsnTrrPheAspAspArgThrTrrPrrPrrLeuGluValArg 80
Db 1433 AATTCGTTTACAGTTTTCAGCTGTTTGGCTTGTGACGACCTAGTGGCTTCATCTGGCAGGC 1492
Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrrPrrHisAla 100
Db 1493 GTTACTTGGAGGACTTATATCTGCTGTTTGTGACGACCTAGTGGCTTCATCTGGCAGGC 1552
Qy 101 LeuArgAsnTrrPrrLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
Db 1553 CTGCGGCACTGTTGGGCTTGGCCATTGACATCAGAACGTTTGTGTCTATTTCGCCCA 1612
Qy 121 LeuPheSerGlyValThrAlaTrrPrrAlaThrTyrGluPheThrLysGluIleLysAspAla 140
Db 1613 CTATTTCTGGGGTCAACCGCTGGGCGACTTACGAAATTTACGAAAGAGATTAAGATGCC 1672
Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
Db 1673 AGCGCTGGGCTTTGGCTGCTGTTTATAGCCATTGTCGCCGTTATATCTAGATCA 1732
Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
Db 1733 GTGGCGGGTCTACGATAATAGGCCATTGGCCATTACACTATTAAATGGTGCACCTTTTCATG 1792
Qy 181 PheTrrPrrLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
Db 1793 TTTTGGATTAGGCCCAAGAACTGGCTCTATCATGACGCAACGCTGTGCGAGCTTTATTC 1852
Qy 201 TyrPheTyrMetValSerAlaTrrPrrGlyGlyTrrValPheIleThrAsnLeuIleProLeu 220
Db 1853 TACTTCTACATGGTGTGGGCTGGGATGATACGTTTTCATCACAACCTTGTATGCCACT 1912
Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
Db 1913 CATGTCTTTTGTGATTTTGTGGGACATATTCCTCAACGTTATCTGCTACACC 1972
Qy 241 ThrTrrPrrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
Db 1973 ACTTGGTACGCTATTGGAACTGTTCATCCATGACGATCCCATTTGCGGTTTCCCTACCT 2032
Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
Db 2033 ATCAGGTCTAACGACACATGGCCGCTTGGGCTGTTTTCGTTTGTGATTTCAGATTGTGCC 2092
Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
Db 2093 TTCGGTGACTCTGTGAAGGCGCAATATCAGCAGCTAAGTTTAAAGTATCATCATGATGTT 2152
Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
Db 2153 TCTCTGTTTGTATCTGCTGCTTGGTGTGGTGGACCTTCTGCTTGGACCTATATGGGG 2212
Qy 321 LeuIleAlaProTrrPrrGlyArgPheTyrSerIleTrrPrrAspThrAsnTrrAlaLysIle 340
Db 2213 TTGATTGCCCTTGGACTGTTAGATTATTTATTCGTTATGGGATACCACTACGCAAGATC 2272
Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrrPrrAlaPhePhe 360
Db 2273 CACATTCTCATCTATCCCTCCGTTCCGAAATCAACCCGTTTCTGTCGCCGCTTTCTTC 2332
Qy 361 PheAspThrHisPheLeuIleTrrPrrPheProAlaGlyValPheLeuLeuPheLeuAsp 380

RESULT 4
AX595714
LOCUS

AX595714 2000 bp DNA linear PAT 14-FEB-2003

Db 2333 TTTGATACCCACATTTTGTGATCTGGCTATTATCCCGCGGTGTATTCTCTACTATTCCTCGAC 2392
Qy 381 LeuLysAspGluHisValPheValIleAlaTrrSerValLeuCysSerTrrPrrPrrPrrPrrPrr 400
Db 2393 TTGAAGACGACGACGCTTTTGTGCTACTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2452
Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
Db 2453 GTTATGTTAGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATTGATT 2512
Qy 421 LeuSerLysIlePheAspIleTrrLeuAspPheLysThrSerAspArgLysTrrAlaIle 440
Db 2513 TTGTCGAAGATATTGACATCTTACCTGGATTTCACAGACAAGTGACGCAATATACGCCATC 2572
Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTrrLeuTrr 460
Db 2573 AAACCTCGGCACCTTACGCGCAAAATGATTGTTTCCGGATCATCTCTTTTATTGTGAT 2632
Qy 461 LeuPheValPheHisSerTrrTrrValThrArgThrAlaTrrSerSerProSerValVal 480
Db 2633 CTTTTCGCTTCCATCTACTTGGGTAAACAAGAACTGCATCTCTTCTCTCTCTCTCTCTCT 2692
Qy 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTrr 500
Db 2693 TTGCGCATCAAAACCCAGATGTTAAATTTGGCGTTTGTGACGACTTCAGGGAAGCGTAC 2752
Qy 501 TyrTrrPrrLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrrPrrPrrPrrPrr 520
Db 2753 TATTGGTTAAGATGAATCTGATGAGGACAGTAGTAAAGTTCAGCGGTGGTGGATACCGT 2812
Qy 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrrPrrAsnAsn 540
Db 2813 TACCAAAATTTGGTGATCGGCAGACAGAACCACTTTAGTCGATACCAACACGCTGGAACAT 2872
Qy 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTrrGluIle 560
Db 2873 ACTCACATCGCATCGTGTGTTAAAGCCATGGCTTCCCTCAAGAGAAATCTTACGAAAT 2932
Qy 561 LeuLysGluHisAspValAspTrrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
Db 2933 CTAAAGACGATGATGCTGATTATGCTTGGTCATCTTTGGTGGTCTAATTTGGGTTTGGT 2992
Qy 581 GlyAspAspIleAsnLysPheLeuTrrPrrMetIleArgIleSerGluGlyIleTrrPrrProGlu 600
Db 2993 GGTGATGACATCAACAAATTTCTTGTGATGATCAGAAATTAGCGAGGAATCTGGCGACAA 3052
Qy 601 GluIleLysGluArgTrrPrrPheTrrThrAlaGluGlyGluTrrArgValAspAlaArgAl 620
Db 3053 CAGATTAAGAGCGTGA-TTTCCTATACCGCAGAGGAGAAATACAGAGTAGATGCAAGGGC 3111
Qy 620 AserGluThrMetArgAsnSerLeuLeuTrrLysMetSerTrrLysAspPheProGlnIle 640
Db 3112 TTCTGAGACCATGAGGAATCTGCTACTTTTACAAAGATGCTCTCAAAAGATTTCCCAAT 3171
Qy 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
Db 3172 ATTCAATGGTGGCCACCCACTGACAGAGTGGCTCAACAAATGATCACCATTAGACGT 3231
Qy 660 lProProLeuAspTrrPrrPheAspGluValPheThrSerGluAsnTrrPrrMetValArgIleTrr 680
Db 3232 CCCACCATTAGACTTCTGCGAAAGTTTACTTCCGAAATCTGATGATGTTAGGATATA 3291
Qy 680 rGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
Db 3292 TCNATTGAGAGAGATGATGCCCAAGGTAGAACTTTTGGGAGCGTTGGTGGTTAACCCAG 3351
Qy 700 gSerSerThrLysThrArgSerIleLysArgProGluLeuGlyLeuArgVal 718
Db 3352 GTCTTCTACGAAACCCAGAGGTCCTAAAGACACCTGAATTAGGCTTGAGAGTC 3406

```

DEFINITION   Sequence 1368 from Patent EP1258494.
ACCESSION    AX595714
VERSION      AX595714.1
KEYWORDS     Saccharomyces cerevisiae (baker's yeast)
SOURCE       Saccharomyces cerevisiae
ORGANISM     Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
              Eukaryota; Fungi; Ascomycota; Saccharomycotina;
REFERENCE    1
AUTHORS      Bauer, A., Gavin, A.C., Grandi, P., Krause, R., Kruse, U., Kuester, B.,
              Mariöch, M., Schult, J., and Superti-Furga, G.
TITLE        Multiprotein complexes from eukaryotes
JOURNAL      Patent: EP 1258494-A 1368 20-NOV-2002;
              CELLZONE AG (DE)
FEATURES     Location/Qualifiers
              1..2000
              /organism="Saccharomyces cerevisiae"
              /mol_type="unassigned DNA"
              /db_xref="taxon:4932"
ORIGIN
Alignment Scores:
Pred. No.:      0      Length:      2000
Score:          605.00  Matches:      665
Percent Similarity: 99.70%  Conservative: 0
Best Local Similarity: 99.70%  Mismatches: 1
Query Match:      84.26%  Indels:      2
DB:               6      Gaps:       0

US-10-028-384-6 (1-718) x AX595714 (1-2000)
QY      1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
DB      1 ATGGGATCCGACCGGCTGCTGTTGTTCTGTTGTTTCAGACCACTCTCAAGCTCGTCATC 60
QY      21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaIleLysPheGlu 40
DB      61 TTCGGTGGCGATTTTGGGGCTGGCCATATCATACAGTTTGTTCAGCTCATCAAAATTGAG 120
QY      41 SerIleIleHisGluPheAspProThrPheAsnThrArgAlaThrLysTyrLeuValAsn 60
DB      121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTATAGGGCTACCAAAATATCTCGTCAAC 180
QY      61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
DB      181 AATTCGTTTACAGTTTTCAGCTGGTTTGACGCTGTTTGACGACCGTACCTGGACCCCTCGGAGG 240
QY      81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
DB      241 GTTACTGGAGGAGCTTTATATCTCTGGTTTGTATGACGACTAGTGGCTTCATCTGGCACGCC 300
QY      101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
DB      301 CTGGCCCACTGGTTGGGCTTGCCATTTGACATCAGAAACGTTTGTTGTTGTTATTTGGCCA 360
QY      121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
DB      361 CTATTTCCTGGGGTCACCGCTGGGCGACTTACGAAATTTACGAAAGAGATTAAAGATGCC 420
QY      141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
DB      421 AGCGCTGGGCTTTGGCTGCTGGTTTATAGCCATATGTCCTCCGGTTATATATCTAGATCA 480
QY      161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
DB      481 GTGGCGGGGCTCTACGATAATAGGCGCAATGGCCATTACACTATTATATGCTCACTTCATG 540
QY      181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
DB      541 TTTTGGATTAAAGGCCCAAGACCTGGCTCTATCATGACGCAACGCTGCACCTTTATTC 600
QY      201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220

```

```

DB      601 TACTTCTACATCGTGTGGCTGGGTGGATACGTTGTCTATCACCACACTTGATCCCACTC 660
QY      221 HisValPheLeuLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
DB      661 CATGTCTTTTCTGCTGATTTTGTATGGGAGATATTCGTCACAACTGTATTCGCTACACC 720
QY      241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
DB      721 ACTTGGTAGCTATTGGAACTGTTGCATCCATGAGATCCCATTTGTCGGTTTCTACCT 780
QY      261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
DB      781 ATCAGGTCTTAACGACCACTGCGGCAATGGGTGTTTTCGGTTTGTATTCAGATTGTCGCC 840
QY      281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
DB      841 TTCGGTGACTTCGTGAGGGCCAAATCAGCACACAGTAAGTTTAAAGTCATCATGATGGTT 900
QY      301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
DB      901 TCTCTGTTTTCATCTGCTGCTGTTGGTGGTGGACTTTCTGCTTACCTATATATGGGG 960
QY      321 LeuIleAlaProThrThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
DB      961 TTGATTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1020
QY      341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
DB      1021 CACATTCCTATCATTCGCTCGCTTTCGAAACATCAACCCGTTTCGGTGGCCGCTTTCTTC 1080
QY      361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
DB      1081 TTTGATACCACTTTTTCATCTGGCTATTCGCCCGCGGTGTATTCCTACTATTCTCTCGAC 1140
QY      381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
DB      1141 TTGAAGAGCAGCAGCGTTTTTGTCTACGCTTACCTCCGTTCTGTGTTCTGACTTTGCCGT 1200
QY      401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
DB      1201 GTTATGGTTAGATTGATGTTGACTTTGACACCACTCATCTGTGTCCGCGCGCGTCGCA 1260
QY      421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
DB      1261 TTGTCGAAAGATTATTGACATCTACTCGATTTCAGACAAAGTACCGCAAAATACGCCATC 1320
QY      441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
DB      1321 AAACCTGGCGCACTACTGCCAAATTTGATTTTCCGGATCATTCATCTTTTATTGTAT 1380
QY      461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
DB      1381 CTTTTCGCTTCCATCTACTTGGGTACAGAACTGACACTACTCTCTCTCTCTCTGTTGTT 1440
QY      481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
DB      1441 TTGCCATCACAAACCCAGATGGTAAATTTGGCGTTGATCGACGACTTCAGGGAAGCGTAC 1500
QY      501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaIleTrpTrpAspTyrGly 520
DB      1501 TATTGGTTAAGATGAATCTCTGATGAGGACAGTAAGGTTGACGGTGGTGGGATTACGGT 1560
QY      521 TyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
DB      1561 TACCAAAATTTGGTCATGGCAGACAGAACCCATTTTGTGATTAACCAACACGTTGGAAAT 1620
QY      541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGlyLysSerTyrGluIle 560
DB      1621 ACTCACATCGCATCTGTTGGTAAAGCCATGGCTTCCCTCGTAAGAGAAATCTTACGAAAT 1680
QY      561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIleGlyPheGly 580
DB      1681 CTAAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

```

QY 581 GlyAspAlaSerMetLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1741 GGTGATGATCATCAACAAATCTTGGATGATCAGAAATTACGAGGAAATCTGCCAGAA 1800
 QY 601 GluIleGlyGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAl 620
 Db 1801 GAGATAAAGAGCGTGA-TTCTATACCGAGGAGGAGATACAGAGTAGATGCAAGGC 1859
 QY 620 aSerGluThrMetArgAsnSerLeuLeuTyrIleMetSerTyrIleValAspPheProGlnLe 640
 Db 1860 TTCTGAGACCATGAGGAATCTGCTACTTTACAGATGCTCTCAAAAGATTTCCACATTT 1919
 QY 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVa 660
 Db 1920 ATTCAATGTGGCCAGCCACTGACAGAGTGGCTCAACAAATGATCACACCATTAGAGT 1979
 QY 660 lProProLeuAspTyrPhe 666
 Db 1980 CCCACCATTAGACTTTC 1998

RESULT 5

AX819898

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-10-028-384-6 (1-718) x AX819898 (1-2000)

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Db 241 GTTACTGGAGGAGCTTTATATCTTGGTTTGNATGACGACTAGTGGCTTCACTGGCAGCC 300
 QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGCCCAACTGGTTGGGCTTCCCAATTGACATCAGAAACGTTTGTGTGCTATTGTGGCCA 360
 QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrIlysGluIleLysAspAla 140
 Db 361 CTATTTCTGGGGTCACCGCTGGCGACATTACGAATTTACGAAAGAGATTAAGATGCC 420
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 421 AGCGCTGGGCTTTTGGCTGGCTTTTATAGCCATTGTCCCGGTATATATCTAGATCA 480
 QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGCGGGCTCTACGATAATGAGCCATTGCCATTACACTATTAAATGCTCACTTTTCATG 540
 QY 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGGCCCAAGACTGGCTCTATCATGCACGCAACGTTGTGAGCTTTATTTC 600
 QY 201 TyrPheTyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTGTGGCTTGGGTGGATACGTGTTCATCACCAACTGTGATCCCACTC 660
 QY 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLysLeuTyrSerAlaTyrThr 240
 Db 661 CATGTCITTTTGGCTGATTTTGGTGGCAGATATTTCGTCAAAGCTGATTTCGTCTACAC 720
 QY 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGTACGCTATTGGACCTTGGATCCATCCATCCAGATCCCATTTGTGGTTCCTACCT 780
 QY 261 IleArgSerAsnAspHisMetAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGCTTAACGACCAACATGGCGCATTTGGGTGTTTTCGGTTTGTATTCAGATTGTGCC 840
 QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TTCGCTGACTTCGTGAAGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGGTT 900
 QY 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 901 TCTCTGTTTTTGCATCTTGGTCTTGGTGGTGGTGGTCTTCTGCTTGGCTTATATGGGG 960
 QY 321 LeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
 Db 961 TTGATGGCCCTTGGACTGGTGTAGATTTTATTCGTTATGGATACCAACTACGCAAGATC 1020
 QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
 Db 1021 CACATTCCTATCATTTGGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080
 QY 361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1081 TTTGATACCACTTTTGTGATTCGCTTATTCGCGCGGGTGTATTCCTACTATTCCTCGAC 1140
 QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 Db 1141 TTGAAAGACGAGACGCTTTTGTGATTCGCTTATTCGCTTATTCGCTTATTCGCTTATTC 1200
 QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 Db 1201 GTTATGGTTAGATTGATTTGACTTTTGGACACCACTCATCTGTGTGTCGCGCGCTCGCA 1260
 QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 Db 1261 TTGTCCAAGATATTGACATCTACCTGGATTTCAAGACCAAGTACCGCAAAATACGCAATC 1320
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460

Db 1321 AAACCTGGCGCACTACTGGCAAAATGATGTTTCCGGATCAATCATCTTTTATTTGAT 1380
 Qy 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTrpSerProSerValVal 480
 Db 1381 CTTTTCGTTCCATCTACTTGGTAAAGAACTGCATCTCTCTCTCTCTCTCTCTCTCT 1440
 Qy 481 LeuProSerGlnThrProAspGlyLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 500
 Db 1441 TTGCCATCACAAACCCAGATGTAATTTGGCTTGATCGACCACTTCAGGGAAGCGTAC 1500
 Qy 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTrpGly 520
 Db 1501 TATTTGGTAAAGATGATCTCTGATGAGACGATGAGGTTCCAGCTGGTGGATACGGT 1560
 Qy 521 TyrGlnLeuGlyMetAlaAspArgThrThrLeuValAspAsnLeuThrTrpAsnLeu 540
 Db 1561 TACCAAAATTTGGTCATGGCAGACAGAACCACTTTAGTCGATTAACACACGTGGAACAT 1620
 Qy 541 ThrHisLeuAlaLeuValGlyLeuAlaMetAlaSerProGluGluLysSerTrpGluLeu 560
 Db 1621 ACTCATCTGGCATCTGTTGGTAAAGCATGGCTTCCCTGACAGAAATCTACGAAAT 1680
 Qy 561 LeuLysGluHisAspValAspTyrValLeuValLeuValLeuPheGlyLeuLeuGlyPheGly 580
 Db 1681 CTAAGAAGAGCATGATGTCGATTAATCTCTGTCATCTTTGGTGTCTAAATGGGTTGGT 1740
 Qy 581 GlyAspAspLeuAlaLeuValGlyLeuTrpMetLeuArgLeuSerGluGlyLeuTrpProGlu 600
 Db 1741 GGTGATGATCATCAACAAATCTCTGATGATGATGATGATGATGATGATGATGATGAT 1800
 Qy 601 GluLeuLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAlaAlaArgAl 620
 Db 1801 GAGATAAAGAGCGTGA-TTCTATACCGCAGAGGGGAGATACAGAGTAGATGCAAGGC 1859
 Qy 620 aSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
 Db 1860 TTCTGAGACCATGAGGAATCGCTACTTTTACAGATGTCCTACAAAGATTTCCACAAAT 1919
 Qy 640 upheAsnGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
 Db 1920 ATTCAATGGTGCCCAAGCCACTGACAGAGTGCCTCAACAAATGATCACCACTTAGAGCT 1979
 Qy 660 lProProLeuAspTyrPhe 666
 Db 1980 CCCACCATAGACTACTTC 1998
 RESULT 6
 AX830928 2000 bp DNA linear PAR 12-DEC-2003
 LOCUS Sequence 1648 from Patent WO03072602.
 DEFINITION AX830928
 ACCESSION AX830928
 VERSION AX830928.1 GI:39839809
 KEYWORDS Saccharomyces cerevisiae (baker's yeast)
 SOURCE Saccharomyces cerevisiae
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1
 Bauer, A., Gavin, A.C., Superti-Furga, G., Kuester, B., Schultz, J.,
 Marzioch, M., Grandi, P., Krause, R., Kruse, U., Merino, A., Bauch, A.,
 Michon, A.M., Leutwein, C. and Rick, J.
 Protein complexes and methods for their use
 Patent: WO 03072602-A 1648 04-SEP-2003;
 CELLZONE AG (DE)
 LOCATION/Qualifiers
 1..2000
 /organism="Saccharomyces cerevisiae"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4932"
 ORIGIN
 FEATURES
 source
 Alignment Scores:
 Pred. No.: 0 Length: 2000

Score: 605.00 Matches: 665
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 84.26% Indels: 2
 DB: Gaps: 0
 US-10-028-384-6 (1-718) x AX830928 (1-2000)
 Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 1 ATGGGATCGACCGGTCGTGTTGTTGCTGTTTTCAGACCATCTCTCAAGCTCGTCATC 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTGCTGGCGAATTTTGGGGCTGCCATATCATCAGCTTGTGTTGAGTCATCAATTTGAG 120
 Qy 41 SerIleHisGluPheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 Db 121 TCTATTATCATGAATTCGACCCCTGGTTCATATTATAGGGCTACCAATATCTCGTCAAC 180
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
 Db 181 AATTCGTTTTCAGAGTTTGAACCTGTTTACGACCGTACCTGTTACCCCTCGGAGG 240
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheIle 100
 Db 241 GTTACTCGAGGACATTTATCTGTTTGTATGACGACTAGTGGCTTCATCTGGCAGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgLeuValCysValLeuPheAlaPro 120
 Db 301 CTGCGCAACTGTTGGGCTTGCCCATTCACATCAGAAACGTTTGTGTGCTATTTCGCGCA 360
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluLeuLysAspAla 140
 Db 361 CTATTTCCTGGGCTCACCGCTGGCGGACTTACGAATTTACGAAGAGATTAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 421 AGCGCTGGGCTTTTGGCTGCTGTTTATAGCAATGTTCCCGGTTATATATCTAGATCA 480
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGCGGGTCTCTAGCATAATGAGCCATTGCCATTACACTATTATTAATGGTCATTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGCCCAAAAGACTGGCTCTATCATGACGCAAGCTGTGCAGCTTTATTC 600
 Qy 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTGTGGGCTTGGGCTGGATACGTTGTTTCATCAACCACTTGATCCCACT 660
 Qy 221 HisValPheLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 661 CATGCTTTTCTGCTGATTTTGTATGGGAGATATTCTGCCAAACCTGATTCTGCTACACC 720
 Qy 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGTGATCGCTATTGGAACTGTTGTCATCCATGAGATCCCATTTGCGGTTTCTTACCT 780
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGTCTAACGACCAACATGCGCGCATTTGGGTGTTTTCGTTTGTATTTCAGATTGCGCC 840
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TCGGTGACTTCGTGAAGGGCCAAATCAGACACCTAAGTTTAAAGTCATCATGTGGTT 900
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 901 TCTCTGTTTTCATCTTGGTCTTGGTGTGGTCTTTCGCGACTTTCGCTTGCCTATATGGGG 960
 Qy 321 LeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340


```

CR382123 17 170001 1753957
Continuation (11 of 18) of CR382123 from base 1000001 (CR382123 Kluyveromyces lactis str

Alignment Scores:
Pred. No.: 4,866-30 Length: 110000
Score: 39,00 Matches: 39
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x CR382123_10 (1-110000)

QY 512 LysValAlaAlaTirAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThr 531
DB 3607 AAAGTTGACAGCTTGGTGGATATGTTATCAATCGGTGTATGGTGTATAGACCACT 3548

QY 532 LeuValAspAsnThrTrpAsnAsnThrHisLeAlaIleValGlyLysAlaMet 550
DB 3547 CTAGTTGATATAACACTTGGATAACACCCACATTCCTATTGTTGGTAAGGCAATG 3491

RESULT 11
AE016903 300029 bp DNA linear PLN 14-APR-2004
LOCUS Eremothecium gossypii chromosome VI, section 5 of 6 of the complete
DEFINITION sequence.
ACCESSION AE016903 AE016819
VERSION AE016903.1 GI:44985021
KEYWORDS
SOURCE Eremothecium gossypii (Ashbya gossypii)
ORGANISM Eremothecium gossypii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Eremothecium.
1 (bases 1 to 300029)
Dietrich,F.S., Voegelii,S., Brachat,S., Lerch,A., Gates,K.,
Steiner,S., Mohr,C., Pohlmann,R., Luedi,P., Choi,S., Wing,R.A.,
Flavier,A., Gaffney,T.D. and Philippsen,P.
The Ashbya gossypii genome as a tool for mapping the ancient
Saccharomyces cerevisiae genome
Science 304 (5668), 304-307 (2004)
15001715
2 (bases 1 to 300029)
Brachat,S., Voegelii,S.E., Dietrich,F.S., Lerch,A., Gaffney,T. and
Philippsen,P.
Direct Submission
Submitted (20-DEC-2002) Applied Microbiology, Biozentrum,
Universitaet Basel, Klingelbergstrasse 50-70, Basel CH-4056,
Switzerland
This is low coverage sequence generated to identify the complete
set of genes and the gene order on this chromosome. Regions of low
quality are identified. Before doing extensive work on any gene
identified here the sequence should be confirmed.
FEATURES
Location/Qualifiers
1..300029
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/strain="ATCC 10895"
/db_xref="ATCC:10895"
/db_xref="taxon:33169"
/chromosome="VI"
/note="synonym: Ashbya gossypii"
1..36701
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2404"
26449..57449
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag1650"
51501..123305
/organism="Eremothecium gossypii"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2318"
100002..157868
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2457"
130220..181064
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2161"
157733..244785
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2545"
244643..300029
/organism="Eremothecium gossypii"
/mol_type="genomic DNA"
/db_xref="taxon:33169"
/clone="BAC clone bag2180"
complement(142..2918)
locus_tag="AFR429C"
complement(142..2918)
locus_tag="AFR429C"
product="AFR429Cp"
complement(142..918)
locus_tag="AFR429C"
/note="synthetic homolog of Saccharomyces cerevisiae
YKL207W"
/codon_start=1
evidence=not experimental
/product="AFR429Cp"
/protein_id="AAS53800.1"
/db_xref="GI:44985022"
translation="MSLGACACAGADLTLSRLKLVLLPISVVMILGVVRYLMILL
TSKYKSVPRATVTENQYISKAIAMLTNGSLHSEAFRMEQEVLAGVLSGKYLAKGK
AAEQNVLAADPNTEALMGAKNLAGYIPQIVIMWVNYFAGFVLMKLPPLNRF
KEMQSGVTADLADARWSSISWIFSMFGLDPVYVNLDFGDEAASQSMQCOQMFVGG
VFGFTPEALMKYANDLTIAHSCDFDIEDRVLMQYAA"
complement(1098..2852)
locus_tag="AFR430C"
complement(1098..2852)
locus_tag="AFR430C"
product="AFR430Cp"
complement(1098..2852)
locus_tag="AFR430C"
/note="NOBY642; no homolog in Saccharomyces cerevisiae"
/codon_start=1
evidence=not experimental
/product="AFR430Cp"
/protein_id="AAS53801.1"
/db_xref="GI:44985023"
translation="MCRLRDAGHRGGAGAFASVPTTKQSVDECPNGLYAKVPHY
RNGAKTPLNLPYQRPKCRFTSASVDKFIREFVALLKQDPLALFENTFENTLDT
TILHWVTSEESRELEASIEAGMGRHRLAKQKQAFETFIIVTDIHAEWLRDSA
WQLSVYQRFIKSDPALREITKGAINTQAEYVKGNYCNAPFPHPHSGVGRSSIDDV
LPLPNWEQFECKYEIDSLASFLTSRQYHENAPDSDRDFITEDWLAAYKIMKLVK
AQMSFTDTGTHVQPTKYKFORSTKIASETPLAGTGNPNVNTGLVRAFRPDDAT
IFQFPVGNIAVLELEHLTAILOSLGDDAAAAPARKVSPSTDLGLLEASQLS
IRFEDGIMEHAYLEHSEKGEVAYEDVGYGSHLPMDANIPLLSIPDLGFGVKDDAV
YRNRKMLTETGPNYNGEHFQGGPHGLHNAPWMSLVRINTSDDEINHCL
DMVKNKTGGGLIHESVRAHPDPGLSYTRSNFANSEFAKTLIDAEKRPHLIFNQE
DAATFPLETFLSQSASV"
1193
/note="region of low quality"
1338
/note="region of low quality"
2809
/note="region of low quality"
2881..2882

```

unsure
unsure
unsure
unsure

ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (sites)
Yoshida, S., Matsura, A., Merregaert, J., and Anraku, Y.
Schizosaccharomycetes pombe att3 is a functional homologue of
Saccharomycetes cerevisiae ST3 which regulates
Oligosaccharyltransferase activity
Yeast 15 (6), 497-505 (1999)
99251095

REFERENCE
MEDLINE 10234787
PUBMED 10234787
2 (bases 1 to 5048)
Yoshida, S.
Direct Submission
Submitted (04-JUN-1998) Satoshi Yoshida, Kirin Brewery Co., Ltd.,
Central Laboratories for Key Technology; 1-13-5, Fukuura
Kanaazawa-Ku, Yokohama, Kanagawa 236-0004, Japan
(E-mail: satoishi@kirin.co.jp, Tel: 81-45-788-7360,
Fax: 81-45-788-4042)

FEATURES
source
1. .5048
/organism="Schizosaccharomycetes pombe"
/mol_type="genomic DNA"
/db_xrefs="taxon:4896"
join(1203..1302,1362..3511)
/gene="att3+"
join(1203..1302,1362..3511)
/gene="att3+"
/note="similar to S. cerevisiae, mouse and human
homologue"
/codon_start=1
/product="Oligosaccharyltransferase subunit"
/protein_id="BAA76479.1"
/db_xref="GI:4586803"
/translation="MANSATITKSGVSKHQKWKIPKVLILICIAVASVSRLLFSV
IYGEIIEFDPNFNRKILVEQFYNFLNFDERSWYPLGRVAGTLYPGIMVTS
GIKFLVLLLRINNRDVCVLLAPFSGITAIATYVARELKSADGLAAAFMGIA
PGVTSRVSAGVDNEAJAITLLMSTFALWIKAVKSGSPWGACGLLYFYMTWAGSY
VETIKMILPHVFLLMGRYTSKLYATITVYVIGTILASQVFPVGFQVSTSEHSA
LGVGLQLFAPFNVGLSVSKQFQILIRFALVCLVGLAVLFAUSSTJLAPWIG
RYSIWDNTAKIHPIIASVSEHPPTWSSFFDLQFLWLLPVGVLGFCRLRNEH
VPIIIVLGTFCGVMLVLTTPCVIAAAVAISTLLDTYMGPEVEDKVSSEA
SAKSNKKGISILFFTSKNGIYISLLSRVISTAYFLIMFYHVSMTSNAY
SSPTVLSVLDGSLMYIDDFRAYDMLRNTPTDKVMSWDYGVQIAGMDRITL
VNNNTWNTIATVCKAMSPEEKAYPILRKHDVYIILYGGTLYGSSDDMKFLWM
IRISGLGDEIVENFTNGEYTDAAPTWRESILKMSVHGAKLFPNQGYD
RRNQKLPKSPQQLTISEATVTHVLVLYKKKPTLGRDLKQVTLFEGRKKSAA
VLQKTKFL"

intron
1303..1361
/gene="att3+"

ORIGIN
Alignment Scores:
Pred. No.: 1,358-23 Length: 5048
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.46% Indels: 0
DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x AB015232 (1-5048)

QY 322 IleAlaProTyrThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHis 341
Db 2246 ATCGCTCTTGACAGGAGTTTCTTCTTTGGGATACAAATACGCCAAGATTTCAT 2305

QY 342 IleProIleAlaSerValSerGluHisGlnPro 353
Db 2306 ATTCTTATCATGCTTGGTATCAGACATCAGCT 2341

RESULT 14
SPBC1271/c

LOCUS SPBC1271 40907 bp DNA linear PLN 23-JUN-2003
S.pombe chromosome II cosmid c1271.
ACCESSION AL034353
VERSION AL034353.2 GI:6912224
KEYWORDS Schizosaccharomycetes pombe (fission yeast)
Schizosaccharomycetes pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (bases 1 to 40907)
Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R.,
Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D.,
Bowman, S., Brooke, K., Brown, D., Brown, S., Chillingworth, T.,
Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P.,
Felwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D.,
Hidalg, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S.,
Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M.,
Leather, S., McDonald, S., McLean, J., Moore, P., Moulé, S.,
Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neill, S.,
Pearson, D., Quail, M.A., Rabinowitsch, E., Rutherford, K., Rutter, S.,
Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M.,
Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G.,
Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J.,
Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weijens, I.,
Vanstraelen, E., Rieger, M., Schaefer, M., Muller-Auer, S., Gabel, C.,
Fuchs, M., Dusterhoft, A., Fritsch, C., Holzer, E., Moestl, D.,
Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H.,
Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H.,
Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S.,
Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z.,
Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gallard, C.,
Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L.,
Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A.,
Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L.,
Lowe, T., McComb, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V.,
Usery, D., Barrell, B. G. and Nurse, P.
The genome sequence of Schizosaccharomycetes pombe
Nature 415 (6874), 871-880 (2002)
21848401
11859360

2 (bases 1 to 40907)
Lyne, M., Rajandream, M.A., Barrell, B.G., Beck, A., Borzym, K.,
Klages, S., Langer, I. and Reinhardt, R.
Direct Submission
Submitted (23-NOV-1998) European Schizosaccharomycetes genome
sequencing project, Sanger Institute, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk and
Max Planck Institut fuer Molekulare Genetik, Ihnestr. 73, D-14195
Berlin, Germany
On Feb 7, 2000 this sequence version replaced gi:3925769.
Notes:
Details of S. pombe sequencing at the Sanger Institute are
available on the World Wide Web
(URL, <http://www.genedb.org/genedb/pombe/index.jsp>) (URL,
http://www.sanger.ac.uk/projects/S_pombe/)
CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
pombe), A (chromosome 1), cSH10 (cosmid name), .01 (first CDS), c
(complementary strand). However, clones may have been reorientated
since the original submission, therefore the complementary strand
notation may be invalid for strand inference. IMPORTANT: This
sequence MAY NOT be the entire insert of the sequenced clone. It
may be shorter because we only sequence overlapping sections once.
Location/Qualifiers
1. .40907
/organism="Schizosaccharomycetes pombe"
/mol_type="genomic DNA"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="II"
/map="IIL"
/clone="cosmid c1271"
1307..3367

FEATURES
source
1. .40907
/organism="Schizosaccharomycetes pombe"
/mol_type="genomic DNA"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="II"
/map="IIL"
/clone="cosmid c1271"
1307..3367

Gene

```

/ gene="SPBC1271.15c"
1307. 3347
/ gene="SPBC1271.15c"
/ note="translation initiation factor if-2 (predicted);
localization mitochondrion (predicted); non-essential
(PMID 12618370); similar to S. cerevisiae YOL023W"
/ codon_start=1
/ protein_id="CAA22205.1"
/ db_xref="GI:3925784"
/ db_xref="GOA:O59683"
/ db_xref="Swiss-Prot:O59683"
/ translation="MARLSACTSRHVNAGTFITKFPFVNHHKLSYHISPLSSRF
SLADISVKNHNPQLRQTHFFHSSPMATKGFREPKSSTLTITPAVLISVASPA
NLQIPTRLDRLNIGIKDVEPYLITYEYSSLAEEYDFDQEAAYGAPSPKSL
NSKSDGSLPVPVLLMGHVDHGKTLTLLDAFKSTIASTEHGGITQKGAFTPFDKG
SKFTFTDPTCHMAFEMRRGRGADIIVLVVAGDGVKQPTVEAIHQISADVPVW
VALTKSRDPTGPIHKIYEOLLNNGIOVEALGGETQIPIISAKTGKIGPELAAITLA
EIMRIRASPRDPFGWIVSSVTKGKGVVVKGTGVMKGMVLVAGKSWCKVRSLV
DVNKSLIKOVLPGQAOVYCKDLPAGDLAYEVKESSEAKEILSDIVRQNEQNFVE
LAESQNRQVSAALAKSGAPAAIQEETSVKSFNIAKCEDTGSSEALSDYKLPQFG
KVRSRVITGVPTVETDIERASTSDAIIISFGVSVFKAFPRIAEKHNVKLLFHNVIY
HLMDDVRKLFALRPPLPILVQRTGEAIIISAFIDIKAKRAVHVAGRVNVTGKSHK
IRLVNDKIIWSEBISLKLKEEVTSIKKRECGILLKNFDEIVTGDKLOTFVBEYK
PPDP"
1811. 2815
/ gene="SPBC1271.15c"
/ note="Pfam match to entry PF00009 GTP_EFTU, Elongation
fac cor Tu family"
1814. 2815
/ gene="SPBC1271.15c"
/ note="Match to GTP_EFTU , Score 114.20"
3215. 3283
/ gene="SPBC1271.15c"
/ note="PS01176 Initiation factor 2 signature"
/ complement(3720.5057)
/ gene="SPBC1271.14"
/ complement(3720.5057)
/ gene="SPBC1271.14"
/ EC number="2.3.1.35"
/ note="amino-acid N-acetyltransferase (predicted);
glutamate N-acetyltransferase (predicted); involved in
ornithine biosynthesis (predicted); involved in arginine
biosynthesis (predicted); amino acid N-acetyltransferase
activity; glutamate N-acetyltransferase activity;
localization mitochondrion matrix (predicted);
non-essential (PMID 12618370); similar to S. cerevisiae
YMR062C
PATH:map00220
aspect:P; term=ornithine biosynthesis; GOid=GO:0006592;
evidence=ISS; db_xref=GOC:unpublished; with=SPTR:Q04728;
date=20030616
aspect:P; term=arginine biosynthesis; GOid=GO:0006526;
evidence=ISS; db_xref=GOC:unpublished; with=SPTR:Q04728;
date=20030616
aspect:P; term=amino acid N-acetyltransferase;
GOid=GO:0004042; evidence=ISS; db_xref=GOC:unpublished;
with=SPTR:Q04728; date=20030616
aspect:P; term=glutamate N-acetyltransferase;
GOid=GO:0004358; evidence=ISS; db_xref=GOC:unpublished;
with=SPTR:Q04728; date=20030616
aspect:C; term=mitochondrial matrix ; GOid=GO:0005759;
evidence=ISS; db_xref=GOC:unpublished; with=SPTR:Q04728;
date=20030616
/ codon_start=1
/ protein_id="CAA22204.1"
/ db_xref="GI:3925783"
/ db_xref="GOA:O94345"
/ db_xref="TrEMBL:O94345"
/ translation="MOTKIFARLAGIGARFVGTMDKTKYFKISSTYDQGSINGIA
SGVXANGKDDIALFSPPCNAAAVFTKNAPQAPVQVRQILNGCGGADHCVFVNS
GCANAVTGGEGMADQILITAEADNLTPHTWSTENSEBFFPSLVNVTGVIQORLKD
KIQSGLEHAVEDLGSSTHEYWRAAEAICTTDTTFPKLVSLRELSIAGKVYRIAGFAGG
MINPNLATLGLFVTDAPISVDVARSILRHAINNSFNSISIGDGTSTNDTIAFLAKGA
AGGSITKSPAYKEIRDVTDIAQKLVVRDGEATKFTVTVVRGARSEKDAALV
ASTIENSALVKTAFPGEDANWGRILCAVYSGAANPPATTVSFIPADGTPEPLKLVN
GSPQNVDSITRASELISQDELTVDDVLCGGPYAKNTWTCDFSYDYVRINADYS"
complement(3723.4946)
/ gene="SPBC1271.14"
/ note="Match to ArgJ , Score 840.54"
/ complement(6265.6931)
/ note="SPMBIRNA D31852 Fission yeast DNA for meirna,
complete sequence"
complement(join(7889.8438,8502.8575))
/ gene="SPBC1271.13"
complement(join(7889.8438,8502.8575))
/ note="mitochondrial ribosomal protein L17; essential
(PMID 12618370); similar to S. cerevisiae YUL063C"
/ codon_start=1
/ protein_id="CAA22203.1"
/ db_xref="GI:3925782"
/ db_xref="GOA:O94345"
/ db_xref="TrEMBL:O94345"
/ translation="MKTLYTRKLRGRTSAHRQALLRLVSLVKHESIQTWAKAEA
QFPAKLIIMAKRANPONRKGAEVMFKEITLKKVDFVLVPRYNGRCGTVRLIK
LPRTDNAPGVLEFVDGPKDIEFHTAKTVGICLAQKALPTITRKNHKKVLLFRK
NGKAEFDQIVQKESEHARLKEDEKTVKDKMGSDPIPRPTV"
complement(8156.8437)
/ gene="SPBC1271.13"
/ note="Match to Ribosomal_L17 , Score 114.95"
/ complement(8439.8454)
/ gene="SPBC1271.13"
/ note="ctaataatgacagacag, splice branch and acceptor"
complement(8496.8501)
/ gene="SPBC1271.13"
/ note="gtagtg, splice donor sequence"
complement(8899.9327)
/ note="rRNA from AB016222"
complement(join(9328.9869,9929.10241,10424.10546,
10690.10768,10941.10966,11042.11054,11105.11175))
/ gene="kes1"
/ note="oxysterol binding protein (predicted); involved in
ergosterol biosynthesis (predicted); similar to S.
cerevisiae YHR001W and YHR003W and YOR237W and YPL145C;
similar to S. pombe SPBC646.08c and SPBC354.07c and
SPCC336.01; localization golgi (predicted); non-essential
(PMID 12618370)"
/ codon_start=1
/ protein_id="CAA22202.1"
/ db_xref="GI:3925781"
/ db_xref="GOA:O74178"
/ db_xref="TrEMBL:O74178"
Alignment Scores:
Pred. No.: 8,93e-23 Length: 40907
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.46% Indels: 0
DB: 8 Gaps: 0
US-10-028-384-6 (1-718) x SPBC1271 (1-40907)
Qy 322 IleAlaProThrGlyArgPheTySerLeuTipAspThrAsnTyAlaLysIleHis 341
Db 36751 ATCGCTCCTTGGACAGACGTTCTTCTTCTTTGGATACAAACTACCCAGATTCAT 36692
Qy 342 IleProIleIleAlaSerValSerGluHisGlnPro 353
Db 36691 ATTCCATCATTCCTTCGGTATCAGAACATCAGCCT 36656

```

RESULT 15

CR382134_03

WPCOMMENT

Sequence split into 14 fragments LOCUS CR382134 Accession CR382134

| Fragment Name | Begin | End |
|---------------|---------|---------|
| CR382134_00 | 1 | 110000 |
| CR382134_01 | 100001 | 210000 |
| CR382134_02 | 200001 | 310000 |
| CR382134_03 | 300001 | 410000 |
| CR382134_04 | 400001 | 510000 |
| CR382134_05 | 500001 | 610000 |
| CR382134_06 | 600001 | 710000 |
| CR382134_07 | 700001 | 810000 |
| CR382134_08 | 800001 | 910000 |
| CR382134_09 | 900001 | 1010000 |
| CR382134_10 | 1000001 | 1110000 |
| CR382134_11 | 1100001 | 1210000 |
| CR382134_12 | 1200001 | 1310000 |
| CR382134_13 | 1300001 | 1349926 |

Continuation (4 of 14) of CR382134 from base 300001 (CR382134 Debaryomyces hansenii chrC

Alignment Scores:

| Pred. No.: | 2.18e-22 | Length: | 110000 |
|------------------------|----------|---------------|--------|
| Score: | 32.00 | Matches: | 32 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 4.46% | Indels: | 0 |
| DB: | 8 | Gaps: | 0 |

US-10-028-384-6 (1-718) x CR382134_03 (1-110000)

| | | | |
|----|-------|--|-------|
| Qy | 322 | IleAlaProThrThrGlyArgPheTyrsSerLeuThrAspThrAsnTyAlaLysIleHis | 341 |
| Db | 65525 | ATTGCACCTTGGACTGGCGTTCTATTCTTTATGGGATCTAACTATGCCAAGATTTCAC | 65584 |
| Qy | 342 | IleProIleIleAlaSerValSerGluHisGlnPro | 353 |
| Db | 65585 | ATTCTTATTAATTCGATCCGTTTCTGACATCAACCT | 65620 |

Search completed: December 14, 2004, 22:41:46

Job time : 8286.53 secs

27/11/2020

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 827.256 Seconds
(without alignments)
4556.130 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 718

Sequence: 1 WGSDRSCVLSVFQITLKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2.1/USPTO spool/US10028384/runat 14122004 131645 6790/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -QFMT=fastap -SUFFIX=Oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn.1.1986@runat 14122004 131645 6790 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -YGAPOP=60 -YGAPEXT=7
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N Geneseq 23Sep04:*

1: geneseqn1980as:.*
2: geneseqn1990as:.*
3: geneseqn2000as:.*
4: geneseqn2001as:.*
5: geneseqn2002as:.*
6: geneseqn2003as:.*
7: geneseqn2004as:.*
8: geneseqn2005as:.*
9: geneseqn2006as:.*
10: geneseqn2007as:.*
11: geneseqn2008as:.*
12: geneseqn2009as:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 718 | 100.0 | 2733 | 10 | ADD94787 Yeast STT |
| 2 | 605 | 84.3 | 2000 | 10 | Acc61293 Gene sequ |
| 3 | 605 | 84.3 | 2000 | 10 | Adk63659 Disease t |
| 4 | 121 | 16.9 | 381 | 6 | Abn23953 Human ORF |
| 5 | 32 | 4.5 | 2256 | 6 | Abz32036 Candida a |
| 6 | 21 | 2.9 | 1848 | 8 | Abt18969 Aspergill |

| | | | | | |
|----|----|-----|------|----|-------------------|
| 7 | 21 | 2.9 | 1969 | 8 | ABT18375 |
| 8 | 21 | 2.9 | 2232 | 8 | ABT20789 |
| 9 | 21 | 2.9 | 2466 | 10 | Abb69900 C. neofo |
| 10 | 21 | 2.9 | 2603 | 8 | ABT20131 |
| 11 | 21 | 2.9 | 2882 | 10 | ADb69539 |
| 12 | 21 | 2.9 | 3969 | 8 | ABT17781 |
| 13 | 21 | 2.9 | 4603 | 8 | ABT19595 |
| 14 | 21 | 2.9 | 4738 | 10 | ADb69178 |
| 15 | 20 | 2.8 | 2417 | 10 | ADD94789 |
| 16 | 20 | 2.8 | 2699 | 4 | ABL13247 |
| 17 | 20 | 2.8 | 2953 | 4 | ABL18224 |
| 18 | 20 | 2.8 | 4922 | 4 | ABL13246 |
| 19 | 17 | 2.4 | 313 | 8 | ABZ20328 |
| 20 | 17 | 2.4 | 387 | 4 | ABA08357 |
| 21 | 17 | 2.4 | 393 | 4 | ABA08547 |
| 22 | 17 | 2.4 | 558 | 5 | AAF93612 |
| 23 | 17 | 2.4 | 1209 | 4 | AAH33264 |
| 24 | 17 | 2.4 | 1209 | 6 | ABL89850 |
| 25 | 17 | 2.4 | 1543 | 2 | AAH85055 |
| 26 | 17 | 2.4 | 1543 | 8 | ACD18981 |
| 27 | 17 | 2.4 | 1543 | 12 | ADG78372 |
| 28 | 17 | 2.4 | 1543 | 12 | ADG78372 |
| 29 | 17 | 2.4 | 1664 | 4 | AAK94184 |
| 30 | 17 | 2.4 | 1664 | 12 | ADL30861 |
| 31 | 17 | 2.4 | 2481 | 10 | ADD94783 |
| 32 | 17 | 2.4 | 2710 | 10 | ADD94785 |
| 33 | 17 | 2.4 | 2855 | 4 | ABL02795 |
| 34 | 17 | 2.4 | 6153 | 4 | ABL02794 |
| 35 | 16 | 2.2 | 154 | 3 | AAC93955 |
| 36 | 16 | 2.2 | 245 | 3 | AAC17015 |
| 37 | 16 | 2.2 | 247 | 2 | AAH40252 |
| 38 | 16 | 2.2 | 474 | 5 | AAF93633 |
| 39 | 16 | 2.2 | 485 | 9 | ACH24071 |
| 40 | 16 | 2.2 | 500 | 6 | ABV88332 |
| 41 | 16 | 2.2 | 1114 | 4 | AAH99794 |
| 42 | 16 | 2.2 | 1371 | 12 | ADP28508 |
| 43 | 16 | 2.2 | 2472 | 10 | ADD94793 |
| 44 | 16 | 2.2 | 2472 | 10 | ADH28844 |
| 45 | 16 | 2.2 | 2760 | 5 | ABV24502 |

ALIGNMENTS

RESULT 1
ADD94787

ID ADD94787 standard; DNA; 2733 BP.

XX

AC ADD94787;

DT 29-JAN-2004 (first entry)

DE Yeast STT3 gene sequence.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Perreault C, McBride K;
WPI; 2003-559122/52.
P-PSDB; ADD94788.

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Disclosure; SEQ ID NO 5; 65pp; English.

This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the yeast SRT3 gene which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

Sequence 2733 BP; 712 A; 597 G; 583 G; 841 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 2733
Score: 718.00 Matches: 718
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADD94787 (1-2733)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrLeuLysLeuValIle 20
Db 219 ATGGGATCCGACCGGTCTGTGTGTGTCTGTGTTCAGACCATCTCTACGCTCGTCATC 278
QY 21 PheValAlaIlePheGlyAlaAlaIleSerArgLeuPheAlaValIleLysPheGlu 40
Db 279 TTCGTGGCGATTTTGGGGCTGCCATATCATCATCGTTTGTTCAGTCATCAATTTGAG 338
QY 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTrpLeuValAsn 60
Db 339 TCTATATTCATGAATTCGACCCCTGGTTCAATTATATAGGCTACCAATATCTCGTCAC 398
QY 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
Db 399 AATTCGTGTTTACAACTTTTGAACCTGGTTTGACGACCGTACCTGGTACCCCTCGGAAGC 458
QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
Db 459 GTTACTGGAGGGACATTTATATCCTGGTTTGATGACGACTAGTCGCTTCATCTGCACGCC 518
QY 101 LeuArgAsnTrpLeuGlyLeuProIleAsnIleArgAsnValCysValLeuPheAlaPro 120
Db 519 CTGCGCACTGTTGGGCTTCCCATTCGACATCAGAAACGTTTGTGTGCTATTGTTCGCCA 578
QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
Db 579 CTATTTTCTGGGGTCCAGCGCTGGCGCACTTACCAATTTACGAAAGAGATTAAAGATGCC 638
QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160

Db 1719 TATTGTTAAGATGAACCTCTGATGAGCAGACAGTAAGGTTGCGAGCTGGTGGGATACGGT 1778
 Qy 521 TyrGlnIleGlyClyMetAlaSerPheThrLeuValAspAsnAenThrTrpAsnAen 540
 Db 1779 TACCAAAATGGTGGCAGTGCAGACAGAACCACTTTAGTCGATACCAACAGCTGGAAACAT 1838
 Qy 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTrpGluIle 560
 Db 1839 ACTCACATCGCCATGTTGTTAAAGCCATGGCTTCCCTGAAGAGAAATCTTACGAAAT 1898
 Qy 561 LeuLysGluHisAspValAspTrpValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1899 CTAAAGAGCATGATGTCATATGTCITGGTCACTTTGTGTGTCTAAATGGGTGGT 1958
 Qy 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1959 GGTGATGATCAACAAATCTTGTGGATGATCAGAAATAGCGAGGAACTCGGCCAGAA 2018
 Qy 601 GluIleLysGluArgTrpPheTrpThrAlaGluGlyGluTrpArgValAspAlaArgAla 620
 Db 2019 GAGTAAAGAGCGCTTATTTCTATCCGAGAGGAGAAATACAGAGTAGATGCAAGGGCT 2078
 Qy 621 SerGluThrMetArgAsnSerLeuLeuTrpLysMetSerTrpLysAspPheProGlnLeu 640
 Db 2079 TCTGAGACCATGAGGAACCTCGTACTTTTACAGATGTCCTACAAAGATTTCCACCAATTA 2138
 Qy 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal 660
 Db 2139 TTCATGTTGGTCCAGCCACTGACAGAGTGGGTCAACAAATGATCACCATTAGAGCTC 2198
 Qy 661 ProProLeuAspTrpPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr 680
 Db 2199 CCACCATTAGACTACTTCGACGAAGTTTCTTACTTCCGAAAACCTGGATGGTTAGAAATAT 2258
 Qy 681 GlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
 Db 2259 CAATTGAAGAGATGATGCCACAGTAGAATTTGAGGACGTTGGTGGTTAACCAAG 2318
 Qy 701 SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2319 TCTTCTACAAAACAGAGGTCCTCAAGAGACCTGAATTAGGCTTTGAGAGTC 2372
 RESULT 2
 ACC61293
 ID ACC61293 standard; DNA; 2000 BP.
 XX
 AC ACC61293;
 XX
 XX 20-JUN-2003 (first entry)
 XX
 DE Gene sequence #SEQ ID 1368.
 XX
 XX Multiprotein complex; eukaryote; drug target; diagnosis; gene; ds.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN EP1258494-A1.
 XX
 PD 20-NOV-2002.
 XX
 XX 20-DEC-2001; 2001EP-00130253.
 XX
 XX 15-MAY-2001; 2001EP-00111774.
 XX
 XX (CELL-) CELLZONE AG.
 XX
 XX Bauer, A, Grandi P, Krause R, Kruse UD, Kuester BD;
 PI Marzioch M, Schultz JD, Superti-Furga GD;
 XX
 DR WPI; 2003-250078/25.
 DR P-PSDB; ABR53251.
 XX
 XX New isolated protein complexes useful for diagnosing a disease or

PT disorder, or as a target for an active agent of a pharmaceutical,
 PT preferably a drug target in the treatment or prevention of disease or
 disorder.
 XX
 XX Disclosure: SEQ ID NO 1368; 17pp + Sequence Listing; English.
 XX
 CC The invention relates to multiprotein complexes from eukaryotes. Proteins
 CC of the invention and DNA sequences encoding them are given in records
 CC ABR52568-ABR53903 and ACC610-ACC61944 respectively. The complexes are
 CC obtainable by using a protein as a bait and isolating the set of proteins
 CC which is attached thereto from cells. Such protein complexes may comprise
 CC up to 30 distinct proteins. Protein complexes of the invention are useful
 CC for diagnosing a disease or disorder, or as a target for an active agent
 CC of a pharmaceutical, preferably a drug target in the treatment or
 CC prevention of a disease or disorder. Note: The sequence data for this
 CC patent is not represented in the printed specification, but is based on
 CC sequence information supplied by the European Patent Office. The complete
 CC document is available on CD-ROM
 XX
 SQ Sequence 2000 BP; 461 A; 459 C; 453 G; 627 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2000
 Score: 605.00 Matches: 665
 Percent Similarity: 99.70% Conservative: 0
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 84.26% Indels: 2
 Ds: 10 Gaps: 0
 US-10-028-384-6 (1-718) x ACC61293 (1-2000)
 Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 1 ATGGGATCCGACCGGCTGTGTGTTTGTCTGTGTTTCAGACCATCTCAAGCTGTCATC 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTCGTGGCGATTTTGGGGCTGCCATCATCATCAGTTTGTGTCAGTCAATCAATTTGAG 120
 Qy 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTrpLeuValAsn 60
 Db 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTTATAGGGCTACCAATATCTCGTCAAC 180
 Qy 61 AsnSerPheTrpLysPheLeuAsnTrpPheAspAspArgThrTrpTrpProLeuGlyArg 80
 Db 181 AATTCGTTTACAGTTTGTGAACCTGGTTGACGACCGTACCTGTACCCCTCGGAAGG 240
 Qy 81 ValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
 Db 241 GTTACTGGAGGACTTTATATCTCTGTTTGTATGACGACTAGTGCCTTCATCTGGCAGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGGCAACTGGTTGGCTTGGCCATGACATCAGAAACGTTTGTGTCTATTTCGCCCA 360
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTrpGluPheThrLysGluIleLysAspAla 140
 Db 361 CTATTTCTGGGCTCACCGCTGGGCGACTTACGAATTTACGAAGAGATTAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTrpLysSerArgSer 160
 Db 421 AGCGCTGGGCTTTTGGCTGCTGTTTATAGCAATGTCCTCCCGTTATATATCTAGATCA 480
 Qy 161 ValAlaGlySerTrpAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGCGGGGCTCTACGATTAATGAGGCCATTGGCATTACACTATTAAATGTCACATTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGCCCAAGAGACTGCTCTATCATGACGACGACCGTGTGACGCTTTATTC 600
 Qy 201 TyrPheTrpMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeu 220

Db 601 TACTTCTACATGGTGTGGCTTGGGTGGATACGTGTTCATCACCACCACTTGATCCCACTC 660
QY 221 HisValPheLeuLeuLeuLeuMetGlyArgTyrSerSerIysLeuTyrSerAlaTyrThr 240
Db 661 CATTGCTTTTGGTGAATTTTGGTGGGAGATATTCGTCCAACTGATTTCTGCCATACC 720
QY 241 ThrTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
Db 721 ACTTGGTACGCTATTGGAACTGTTGTCATCCATGCAATCCCATTTGCGTTTCTTACCT 780
QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
Db 781 ATCAGGCTTACAGACCATGCGCCGATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 840
QY 281 PheGlyAspPheValIleGlyGlnIleSerThrAlaIlePheIysValIleMetMetVal 300
Db 841 TTGGTGTGACTTCTGGTGAAGGCGCAATCAGCACAGCTAAAGTTAAAGTCATCATGATGTT 900
QY 301 SerLeuPheLeuLeuLeuValIleGlyValIleGlyLeuSerAlaLeuThrTyrMetGly 320
Db 901 TCTCTGTTTGTGATCTGGTCCCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 960
QY 321 LeuIleAlaProTyrThrGlyArgPheTyrSerLeuTyrAspThrAsnTyrAlaIle 340
Db 961 TTGATTTGCCCTTGGACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1020
QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTyrProAlaPhePhe 360
Db 1021 CACATTTCTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 361 PheAspThrHisPheLeuIleTyrLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
Db 1081 TTGTGATACCACTTTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 381 LeuIysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
Db 1141 TTGAAAGACGACGACGCTTTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaIleValAla 420
Db 1201 GTTATGGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheIysThrSerAspArgIysTyrAlaIle 440
Db 1261 TTGTCCAAAGATTTTGACATCTACCTGGAATTCAGACAGATGACCGCAATATCCCATC 1320
QY 441 LysProAlaIleLeuLeuAlaIleValIleValSerGlySerPheIlePheTyrLeuTyr 460
Db 1321 AAACCTGCGCACTACTGCGCAATGATTTGCTGCGATCATCTCTTTTATTGTGAT 1380
QY 461 LeuPheValPheHisSerThrTyrValArgThrAlaTyrSerSerProSerValVal 480
Db 1381 CTTTTCGTCTTCCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1440
QY 481 LeuProSerGlnThrProAspGlyValLeuAlaIleValSerPheArgGluAlaTyr 500
Db 1441 TTGCCATCACAAACCCAGATGGTAAATTTGGCTTGTATGACGACTTCAGGAGCGATAC 1500
QY 501 TyrTyrLeuArgMetAsnSerAspGluAspSerIysValAlaIleTyrPheTyrGly 520
Db 1501 TATTTGGTTAAGATGAATCTGTATGAGGACAGTAAGGTTGACGCGTGGTGGATACCGT 1560
QY 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrThrPheAsn 540
Db 1561 TACCAAAATGGTGGCATGGCAGACAGAACCATTTAGTCGATACACACGTTGGACAT 1620
QY 541 ThrHisIleAlaIleValGlyValAlaMetAlaSerProGluIleLysSerTyrGluIle 560
Db 1621 ACTCATCTGGCATCTGTTGGTAAAGCCATGCTTCCCTTGAAGAGAAATCTTACCAAAAT 1680
QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
Db 1681 CTAAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1740

QY 581 GlyAspAspIleAsnLysPheLeuTyrMetIleArgIleSerGluGlyIleTyrProGlu 600
Db 1741 GGTGATGATCATCAAAATTTCTTGTGATGATCAGATTTAGCGGGGATCTGGCAGAA 1800
QY 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAlaArgAl 620
Db 1801 GAGATAAAGAGCGTGA-TTTCTATACCGCAGAGGAGATACAGAGTAGATGCAAGGCG 1859
QY 620 aserGluThrMetArgAsnSerLeuTyrLysMetSerTyrLysAspPheProGlnIle 640
Db 1860 TTCTGAGACCATGAGGAACCTGCTACTTTACAGAGTGTCTCAAGAGATTTCACCAAT 1919
QY 640 uPheAsnGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
Db 1920 ATTCAATGTTGGCCAAAGCCACTGACAGTGCCTCAACAAATGATCACACCATTAGAG 1979
QY 660 lProProLeuAspTyrPhe 666
Db 1980 CCCACCATTAGACTTCTTC 1998

RESULT 3

ADK63659
ID ADK63659 standard; DNA; 2000 BP.

XX AC ADK63659;

XX XX 06-MAY-2004 (first entry)

DT XX Disease treating protein complex-derived gene #824.

DE XX protein complex; drug target; diagnosis; gene; ds.

XX KM Unidentified.

XX OS BP1338608-A2.

XX PN 27-AUG-2003.

XX PD 20-DEC-2002; 2002EP-00102902.

XX PF 20-DEC-2001; 2001EP-00130253.

XX PR (CELL-) CELLZOME AG.

XX PA Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;

XX PI Marzioch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;

XX PI Michon A, Leutwein C, Rick J;

XX DR WPI; 2003-638460/61.

XX DR P-PSDB; ADK63658.

XX PT New proteins and protein complexes from eukaryotes, useful as targets in

PT drug screening, or in diagnosing or screening for the presence of a

PT disease or disorder, or a predisposition for developing a disease or

PT disorder in a subject.

XX PS Disclosure; SEQ ID NO 1648; 13pp; English.

XX CC The invention relates to novel protein complexes comprising a first and a
CC second protein, or its derivative, fragment, homologue or variant. The
CC proteins are selected from given protein complexes, which are not defined
CC in the specification. The variants are encoded by nucleic acids that
CC hybridize to the nucleic acids encoding the proteins under low stringency
CC conditions. The protein complexes are useful as targets for an active
CC agent of a pharmaceutical. These protein complexes are particularly
CC useful as drugs targets for the treatment or preventing of a disease or
CC disorder. The complexes and methods above are useful in diagnosing or
CC screening for the presence of a disease or disorder or a predisposition
CC for developing a disease or disorder in a subject. These are also useful
CC in screening for a drug for treatment or prevention of a disease or
CC disorder. The molecule that modulates the amount, activity or protein
CC components of the complex is useful for the manufacture of a medicament

CC for the treatment or prevention of a disease or disorder. This sequence
 CC corresponds to a gene of the invention. (Note: the sequence data for this
 CC patent did not form part of the printed specification but was obtained
 CC from the EPO in electronic format).

XX
 SQ Sequence 2000 BP; 461 A; 459 C; 453 G; 627 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2000
 Score: 605.00 Matches: 665
 Percent Similarity: 99.70% Conservative: 1
 Best Local Similarity: 99.70% Mismatches: 0
 Query Match: 84.26% Indels: 2
 DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADK63659 (1-2000)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 1 ATGGGATCGACCGGCTGGTGTGTTGCTGTGTTTCAGACCATCTCAAGCTCGTCATC 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTCGGGGCGATTTTGGGGCTGCCATATCATCAGCTTGTGTTGTCAGTCATCAATTTGAG 120
 Qy 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTyrLeuValAsn 60
 Db 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTATAGGCTACCAATAATCTCGTCAAC 180
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgTrpTrpTyrProLeuGlyArg 80
 Db 181 AATTGGTTTTCAGAGTTTTCAGCTGGTTTTCAGACCGCTACCTGGTACCCCTCGGAAGG 240
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
 Db 241 GTTACTGGAGGACATTTATCTGTTTGTATGACGACTAGTGGTTCATCTGGCAGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGGCACTGGTGGGCTTGGCCATGACATCAGAACGTTTGTGCTGTTATTTGGCCCA 360
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 Db 361 CTATTTCTGGGGTCAACCGCTGGGCGACTTACGAATTTACGAAGAAGATTAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 421 ACGGCTGGGCTTTTGGCTGGTGTATAGCCATTGTCCTGGTATATATCTAGATCA 480
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuMetValThrPheMet 180
 Db 481 GTGGGGGGTCTACGATAATGAGGCCATTGCCATTACATTTAATGGTCACTTTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMethHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAGGCCCAAAAGACTGGCTCTATCATGTCAGCAACGTCGTGAGCTTTATTC 600
 Qy 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTGTGGCTGGGGTGGATACGGTTCATCAACCACTTGTATCCCACTC 660
 Qy 221 HisValPheLeuLeuIleLeuMetClyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 661 CATGCTTTTGTCTGATTTTGTATGGGCGAGATATTCTGTCCTCAAACTGATTTCTGCTACACC 720
 Qy 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGGTACGCTATTATGGAACTGTTGTCATCCATGTCAGATCCCATTTGTCGGTTCCTACCT 780
 Qy 261 IleArgSerAspAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGTCTACGACCAATGCGCGCATGGGGTGTGTTTCGGTTTGTATTCAGATTGTCGCC 840

Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TTCGGTACCTTCGTAAGGCCCAAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGTT 900
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaIleThrTyrMetGly 320
 Db 901 TCTCTGTTTTTGATCTTGGTCCCTGGTGGTGGCTTCTGCTGTGACCTATATATGGGG 960
 Qy 321 LeuIleAlaProTrpTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
 Db 961 TTGATTGGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACATACGCAAGATC 1020
 Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
 Db 1021 CACATTCCTATCATTCCTCCGTTTCGAAACATCAACCCGTTTCGGGGCCGCTTCTTC 1080
 Qy 361 PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1081 TTTGATACCCACTTTTGTGATCTGGCTATTCGCCGCGGTGATATTCCTACTATTCTCTGAC 1140
 Qy 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 Db 1141 TTGAAAGACGACGACGTTTGTATCGCTTACTCCGTTCTGTTCTGTTCTGTTCTGTTCT 1200
 Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 Db 1201 GTTATGGTTAGATTGATTGATCTTTGACACACAGTCATCTGTGTGTCGCCGCGTCGCA 1260
 Qy 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 Db 1261 TTGTCCCAAGATATTGTGATCTTACCTGGATTTCAAGACAAAGTAGCCGCAAAATACGCCATC 1320
 Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 Db 1321 AAACCTCGGCACCTACTTGGCCAAATGATTGTTTCCGGATCATTCATCTTTATTTGAT 1380
 Qy 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
 Db 1381 CTTTTCGTTCTCCATCTTACTTGGGTAAAGAACTGTCATCTCTCTCTCTCTCTCTCTCT 1440
 Qy 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
 Db 1441 TTGGCCATCACAAACCCAGATGGTAAATGGCTTGCCTGATCGACACTTACGGAAGCGTAC 1500
 Qy 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGly 520
 Db 1501 TATTGGTTAAGATGAACCTCTGATGAGGACAGTAAGCTTGCACGCTGGTGGGATTACGGT 1560
 Qy 521 TyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 Db 1561 TACCAATTTGGTGCATGCGCAGACAGAACCACTTTAGTCGATTAACACACGTCGAACAAT 1620
 Qy 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluLysSerTyrGluIle 560
 Db 1621 ACTCACATCGGCATCTGTTGTAAGCCATGCTTCCCTCGAAGAGAAATCTTACGAATTT 1680
 Qy 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIleGlyPheGly 580
 Db 1681 CTAAAGAGCATGATGCTGATTAATGCTTGGTCTCATCTTGGTGTCTAATTTGGTGTGCT 1740
 Qy 581 GlyAspAspIleLeuLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1741 GGTGATGACATCAACAAATCTTGTGATGATCGAATTTAGCGGGGAATCTGGCCAGAA 1800
 Qy 601 GluIleLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAl 620
 Db 1801 GAGATAAAGAGCGTGA-TTTTATATACCGCAGAGGGAGAAATACAGATAGATGCAAGGSC 1859
 Qy 620 aSerGluThrMetArgHisSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
 Db 1860 TTCTGAGACCATGAGGAACCTGCTACTTTACAGATGCTCTACAAAGATTTCCCAAT 1919
 Qy 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660

DB 1920 ATTCATGGTGGCCAGCCACTACAGAGTGGCTCACAATAATGATCACACCATAGAGCT 1979
 QY 660 lProPheAspTyrPhe 666
 DB 1980 CCCACCATAGACTACTC 1998
 RESULT 4
 ABN23953
 ID ABN23953 standard; cDNA; 381 BP.
 XX
 AC ABN23953;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX polynucleotide sequence SEQ ID NO:16383.
 KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US010836.
 XX
 PR 30-MAY-2000; 2000US-0206132P.
 PR 29-AUG-2000; 2000US-0228715P.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shimketa RA, Leach MD;
 XX
 DR WPI; 2002-106308/14.
 DR P-PSDB; ABP08201.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders.
 XX
 PS Disclosure; SEQ ID NO 16383; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-1491 (see table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic

CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 381 BP; 100 A; 78 C; 86 G; 116 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 7.03e-115 Length: 381
 Score: 121.00 Matches: 121
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 16.85% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-6 (1-718) x ABN23953 (1-381)
 QY 452 SerGlySerPhePhePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArg 471
 DB 1 TCCGGATCATTCATCTTTTATTTATTTGTTATCTTTTCTTCCATCTCTTGGGTAAACAAGA 60
 QY 472 ThrAlaTyrSerSerProSerValValLeuProSerClnThrProAspGlyLysLeuAla 491
 DB 61 ACTGCATATCT 120
 QY 492 LeuIleAspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSer 511
 DB 121 TTGATCGACGACTTCAGGGAAGCGTACTATTGTTTAAAGATGAATCTCTGATGAGGACAGT 180
 QY 512 LysValAlaAlaTyrTrpAspTyrGlyTyrClnIleGlyGlyMetAlaAspArgThrThr 531
 DB 181 AAGGTTGACGGGTGGTGGGATTACGGTTTACCAATTTGTTGGCATGGCAGACAAACCACT 240
 QY 532 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAla 551
 DB 241 TTAGTCGATAACAACACACGTCGAAACAATACTCACATCCCATCGTTGGTAAAGCCATGGCT 300
 QY 552 SerProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuVal 571
 DB 301 TCCCTGAAGAGAAATCTTACGAAATCTTAAAGAGCATGATGTCGATATGTTCTTGGTTC 360
 QY 572 Ile 572
 DB 361 ATC 363
 RESULT 5
 ABN232036
 ID ABN232036 standard; DNA; 2256 BP.
 XX
 AC ABN232036;
 XX
 DT 30-JAN-2003 (first entry)
 XX
 DE Candida albicans essential gene SEQ ID NO 6323.
 XX
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 XX
 PD 11-JUL-2002.
 XX
 PF 26-DEC-2001; 2001WO-US049486.
 XX
 PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00792024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 DR WPI; 2002-566694/50.

DR P-PSDB; ABP73486.
 XX Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX Claim 37; SEQ ID NO 6323; 167bp + Sequence Listing; English.
 XX The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter-replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX
 XX Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 2,658-22 Length: 2256
 Score: 32.00 Matches: 32
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 4.46% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-6 (1-718) x AB232036 (1-2256)
 QY 322 IleAlaProThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHis 341
 Db 991 ATTGCTCCTTGACACAGGTAGATTTTATTCCTATGGGATACAAATATGCGAGATTCTAT 1050
 QY 342 IleProIleIleAlaSerValSerGluHisGlnPro 353
 Db 1051 ATTCCAATATTGCTTCGTCTTCTGACATCAACCT 1086
 RESULT 6
 ABT18969
 ID ABT18969 standard; DNA; 1848 BP.
 XX
 AC ABT18969;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Aspergillus fumigatus essential gene #1327.
 XX
 KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 OS Aspergillus fumigatus.
 XX
 PN WO200286090-A2.
 XX
 PD 31-OCT-2002.
 XX

PF 23-APR-2002; 2002WO-US013142.
 XX
 PR 23-APR-2001; 2001US-0285697P.
 PR 27-APR-2001; 2001US-0287086P.
 PR 05-JUN-2001; 2001US-0295890P.
 PR 09-JUL-2001; 2001US-0303899P.
 PR 31-AUG-2001; 2001US-0316362P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.
 XX
 PT New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 PS Disclosure; Page; 175pp; English.
 XX
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX
 SQ Sequence 1848 BP; 363 A; 510 C; 447 G; 528 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4,91e-11 Length: 1848
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0
 US-10-028-384-6 (1-718) x ABT18969 (1-1848)
 QY 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 448 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGAACCTGATCCCTGCACGTT 507
 QY 223 Phe 223
 Db 508 TTT 510
 RESULT 7
 ABT18375
 ID ABT18375 standard; DNA; 1969 BP.
 XX
 AC ABT18375;
 XX

DT XX 16-APR-2003 (first entry)
DE XX Aspergillus fumigatus essential gene #733.
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX cancer; contamination; biofilm; antibody; immune response; ds.
OS Aspergillus fumigatus.
XX WO200286090-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US013142.
XX 23-APR-2001; 2001US-0285697P.
XX 27-APR-2001; 2001US-0287066P.
XX 05-JUN-2001; 2001US-0295890P.
XX 09-JUL-2001; 2001US-0303899P.
XX 31-AUG-2001; 2001US-0316362P.
XX (ELIT-) ELITRA PHARM INC.
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX Disclosure; Page; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or contain contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterization, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This polynucleotide sequence represents one of the essential
XX genes of Aspergillus fumigatus of the invention
SQ Sequence 1969 BP; 389 A; 540 C; 480 G; 560 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 5,22e-11 Length: 1969
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 8 Gaps: 0
US-10-028-384-6 (1-718) x ABT18375 (1-1969)
QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuLeuProLeuHisVal 222

Db 514 TACATGGTGTGGCATGGGTGGGTATGCTTTCATTACGAACCTGATCCCTGCACGTT 573
QY 223 Phe 223
Db 574 TTT 576
RESULT 8
ABT20789
ID ABT20789 standard; DNA; 2232 BP.
XX ABT20789;
XX 16-APR-2003 (first entry)
XX Aspergillus fumigatus essential gene #3147.
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX cancer; contamination; biofilm; antibody; immune response; ds.
OS Aspergillus fumigatus.
XX WO200286090-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US013142.
XX 23-APR-2001; 2001US-0285697P.
XX 27-APR-2001; 2001US-0287066P.
XX 05-JUN-2001; 2001US-0295890P.
XX 09-JUL-2001; 2001US-0303899P.
XX 31-AUG-2001; 2001US-0316362P.
XX (ELIT-) ELITRA PHARM INC.
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX Disclosure; Page; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or contain contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterization, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This polynucleotide sequence represents one of the essential
XX genes of Aspergillus fumigatus of the invention

SQ Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.88e-11 Length: 2232
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ABT20789 (1-2232)

QY 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 625 TACATGGTGGCATGGGTGGGTATGTTCTTATTACGAACCTGATCCCTCCGACGTT 684

QY 223 Phe 223
 DB 685 TTT 687

RESULT 9

ADB69900
 ID ADB69900 standard; DNA; 2456 BP.

AC ADB69900;

DT 04-DEC-2003 (first entry)

DE C. neoformans open reading frame SEQ ID NO:2305.

KW ds; gene; fungicide; gene therapy; infection.

OS Cryptococcus neoformans.

PN WO2003052076-A2.

PD 26-JUN-2003.

PF 17-DEC-2002; 2002WO-US040225.

PR 17-DEC-2001; 2001US-0341261P.

PA (ELIT-) ELITRA PHARM INC.

PI Zamudio C, Eroshkin AM;

DR WPI; 2003-533017/50.

DR P-PSDB; ADB70261.

PT New nucleic acid, useful for preparing a composition for treating an infection caused by *Cryptococcus neoformans*.

PS Claim 2; SEQ ID NO 2305; 136pp; English.

CC The invention relates to a novel purified or isolated *Cryptococcus neoformans* nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by *Cryptococcus neoformans*. The present sequence represents a *C. neoformans* sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 2456 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.47e-11 Length: 2456
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0

DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADB69900 (1-2456)

QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 DB 637 ATTGTACCGGATACATCTCTCGATCTGTGCGCGGTCTTATGACACCAAGCCATTGCC 696

QY 172 Ile 172
 DB 697 ATC 699

RESULT 10

ABT20191

ID ABT20191 standard; DNA; 2603 BP.

XX AC ABT20191;

DT 16-APR-2003 (first entry)

DE *Aspergillus fumigatus* essential gene #2549.

KW Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection; cancer; contamination; biofilm; antibody; immune response; ds.

OS *Aspergillus fumigatus*.

PN WO200286090-A2.

PD 31-OCT-2002.

PF 23-APR-2002; 2002WO-US013142.

PR 23-APR-2001; 2001US-0285697P.

PR 27-APR-2001; 2001US-0287066P.

PR 05-JUN-2001; 2001US-0295890P.

PR 09-JUL-2001; 2001US-0303899P.

PR 31-AUG-2001; 2001US-0316362P.

PA (ELIT-) ELITRA PHARM INC.

PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

WPI; 2003-093124/08.

PT New purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*, useful for treating or preventing infections by *A. fumigatus*, or for treating a non-infectious disease in a subject e.g. cancer.

PS Disclosure; Page: 175pp; English.

CC The invention relates to novel purified or isolated nucleic acids of essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as *A. fumigatus*, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by *A. fumigatus*, or to prevent or inhibit formation on a surface of a biofilm comprising *A. fumigatus*. The polynucleotides are useful for expressing recombinant protein for characterization, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of *A. fumigatus* to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for

CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention

XX Sequence 2603 BP; 565 A; 698 C; 620 G; 720 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6,81e-11 Length: 2603
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x ADB20191 (1-2603)

QY 203 TyrMetValSerIaTrpGlyGlyTyrValPheIleThrAsnLeuIleProIleHisVal 222
 DB 877 TACATGGTGTGGCATGGGTGGGTATGTTTCATTACGAACCTGATCCCTGCACGTT 936

QY 223 Phe 223
 DB 937 TTT 939

RESULT 11

ADB69539
 ID ADB69539 standard; DNA; 2882 BP.

XX ADB69539;

XX 04-DEC-2003 (first entry)

DE C. neoformans coding sequence with introns SEQ ID NO:1305.

XX ds; gene; fungicide; gene therapy; infection.

XX Cryptococcus neoformans.

XX WO2003052076-A2.

XX 26-JUN-2003.

XX 17-DEC-2002; 2002WO-US040225.

XX 17-DEC-2001; 2001US-0341261P.

XX (ELIT-) ELITRA PHARM INC.

XX Zamudio C, Eroshkin AM;

XX WPI; 2003-533017/50.

XX P-PSDB; ADB70261.

XX New nucleic acid, useful for preparing a composition for treating an
 PT infection caused by *Cryptococcus neoformans*.

XX Claim 3; SEQ ID NO 1305; 136pp; English.

XX The invention relates to a novel purified or isolated *Cryptococcus*
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by *Cryptococcus neoformans*.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX SQ Sequence 2882 BP; 615 A; 759 C; 668 G; 840 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,51e-11 Length: 2882
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADB69539 (1-2882)

QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 DB 796 ATTGTACCCGGATACATCTCTCGATCTGTCCCGGTTCTTATGACACGAAGCCATTGCC 855
 QY 172 Ile 172
 DB 856 ATC 858

RESULT 12

ABT17781

ID ABT17781 standard; DNA; 3969 BP.

XX ABT17781;

XX 16-APR-2003 (first entry)

DE *Aspergillus fumigatus* essential gene #139.

XX Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.

XX *Aspergillus fumigatus*.

XX WO200286090-A2.

XX 31-OCT-2002.

XX 23-APR-2002; 2002WO-US013142.

XX 23-APR-2001; 2001US-0285697P.

XX 27-APR-2001; 2001US-0287066P.

XX 05-JUN-2001; 2001US-0285890P.

XX 09-JUL-2001; 2001US-0303899P.

XX 31-AUG-2001; 2001US-0316362P.

XX (ELIT-) ELITRA PHARM INC.

XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT *fumigatus*, useful for treating or preventing infections by *A. fumigatus*,
 PT or for treating a non-infectious disease in a subject e.g. cancer.

XX Disclosure; Page; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune

CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX SQ Sequence 3969 BP; 964 A; 996 C; 953 G; 1056 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.02e-10 Length: 3969
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0
 US-10-028-384-6 (1-718) x ABT17781 (1-3969)
 QY 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 1514 TACATGGTGTGGCATGGGTGGGTATGTTCTTACAGAACCTGATCCCTCGACGTT 1573
 QY 223 Phe 223
 DB 1574 TTT 1576
 RESULT 13
 ABT19595
 ID ABT19595 standard; DNA; 4503 BP.
 XX AC ABT19595;
 XX DT 16-APR-2003 (first entry)
 XX DE *Aspergillus fumigatus* essential gene #1953.
 XX KW Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection;
 XX KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX OS *Aspergillus fumigatus*.
 XX PN WO200286090-A2.
 XX PD 31-OCT-2002.
 XX PF 23-APR-2002; 2002WO-US013142.
 XX PR 23-APR-2001; 2001US-0285697P.
 XX PR 27-APR-2001; 2001US-0287066P.
 XX PR 05-JUN-2001; 2001US-0295890P.
 XX PR 09-JUL-2001; 2001US-0303899P.
 XX PR 31-AUG-2001; 2001US-0316362P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WIPI; 2003-093124/08.
 XX DR New purified or isolated nucleic acids of essential genes of *Aspergillus*
 PT *fumigatus*, useful for treating or preventing infections by *A. fumigatus*,
 PT for treating a non-infectious disease in a subject e.g. cancer.
 XX Disclosure; Page; 175pp; English.
 CC The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a

CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*
 CC *fumigatus* to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX SQ Sequence 4603 BP; 1121 A; 1129 C; 1133 G; 1220 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.17e-10 Length: 4503
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 8 Gaps: 0
 US-10-028-384-6 (1-718) x ABT19595 (1-4503)

QY 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 1877 TACATGGTGTGGCATGGGTGGGTATGTTCTTACAGAACCTGATCCCTCGACGTT 1936
 QY 223 Phe 223
 DB 1937 TTT 1939
 RESULT 14
 ADB69178
 ID ADB69178 standard; DNA; 4738 BP.
 XX AC ADB69178;
 XX DT 04-DEC-2003 (first entry)
 XX DE C. neoformans genomic DNA sequence SEQ ID NO:305.
 XX KW ds; gene; fungicide; gene therapy; infection.
 XX OS *Cryptococcus neoformans*.
 XX PN WO20003052076-A2.
 XX PD 26-JUN-2003.
 XX PF 17-DEC-2002; 2002WO-US040225.
 XX PR 17-DEC-2001; 2001US-0341261P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Zamudio C, Eroshkin AM;
 XX WIPI; 2003-533017/50.
 XX DR P-PADB; ADB70261.
 XX PT New nucleic acid, useful for preparing a composition for treating an

PT infection caused by Cryptococcus neoformans.

PS Claim 3; SEQ ID NO 305; 136pp; English.

XX The invention relates to a novel purified or isolated Cryptococcus
CC neoformans nucleic acid molecule comprising a sequence encoding a
CC polypeptide comprising a sequence not given in the specification. A
CC polynucleotide of the invention has fungicide activity, and may have a
CC use in gene therapy. The nucleic acid is useful for preparing a
CC composition for treating an infection caused by Cryptococcus neoformans.
CC The present sequence represents a C. neoformans sequence of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 4738 BP; 1162 A; 1160 C; 1070 G; 1346 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.21e-10 Length: 4738
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADB69178 (1-4738)

Qy 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAla 171

Db 1652 ATTGTACCGGATACATCTCTCGATCTGTCGGGTTCTTATGACAGCAAGCCATTGCC 1711

Qy 172 Ile 172

Db 1712 ATC 1714

RESULT 15

ID ADD94789 standard; DNA; 2417 BP.

AC ADD94789;

XX 29-JAN-2004 (first entry)

DE Drosophila melanogaster STT3 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide;
KW fruit fly; gene; ds; STT3.

OS Drosophila melanogaster.

XX WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX Perreault C, McBride K;

XX WPI; 2003-559122/52.

XX P-PSDE; ADD94790.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung

PT or breast cancer, or for suppressing an immune response in an autoimmune
PT disease.

PS Claim 6; SEQ ID NO 7; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,
CC termed source of immunodominant major histocompatibility complex (MHC)-
CC associated peptide (SIMP), which is expressed ubiquitously in human
CC cells, where the protein has the potential of generating several protein
CC fragments binding with high affinity to a human leukocyte antigen (HLA)
CC molecule. The invention may allow development of therapeutics with
CC cytostatic or immunosuppressive activity or provide sequences useful for
CC antisense therapy or gene therapy. The source of immunodominant MHC-
CC associated peptide (SIMP) nucleic acids, proteins and fragments are
CC useful for diagnosing and treating cancers, for example lung cancer,
CC cancer, melanomas, sarcomas, prostate cancer, testicular cancer, breast
CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
CC proteins are also useful for modulating an immune response. Decreasing
CC lymphoid cell proliferation is useful for suppressing an immune response.
CC responsible for an autoimmune disease or a transplant rejection. The
CC present sequence is that of the Drosophila melanogaster STT3 gene which
CC is related to the invention. Note: This sequence does not appear in the
CC specification but was obtained by the indexer from GenBank.

XX Sequence 2417 BP; 525 A; 628 C; 654 G; 510 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.83e-10 Length: 2417
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADD94789 (1-2417)

Qy 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 521 TTTCATCGCATCGTGGCTGGCTACATCAGTAGTGGTGGCTGGCTAGCTAGTACGAG 580

Search completed: December 14, 2004, 23:34:35

Job time : 856.256 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 139.895 Seconds
(without alignments)
3648.073 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 718
Sequence: 1 WGSRSCLSVFQTLKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table: OLIGO
Xgapop 60.0, Xgapext 60.0
Fgapop 60.0, Fgapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1
Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spoci/US10028384/runat_14122004_131646_6816/app.query.fasta_1.3740
-DB=Issued Patents NA -OFF=fastap -SUFFIX=Oligo.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Oligo -TRANS=human40.cci
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US10028384 @CNG 1 1 259 @runat_14122004_131646_6816 -NCPU=6 -ICPU=3
-NO MVAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=30 -DELOF=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1 | 631 | 87.9 | 2157 | US-09-614-221A-318 | Sequence 318, App |
| 2 | 30 | 4.2 | 857 | US-09-248-796A-3089 | Sequence 3089, App |
| 3 | 26 | 3.6 | 594 | US-09-248-796A-3090 | Sequence 3090, App |
| 4 | 20 | 2.8 | 900 | US-09-270-767-1231 | Sequence 1231, A |
| 5 | 17 | 2.4 | 1365 | US-09-270-767-11648 | Sequence 11648, A |
| 6 | 16 | 2.2 | 245 | US-09-513-999C-21090 | Sequence 21090, A |
| 7 | 16 | 2.2 | 580 | US-09-270-767-1318 | Sequence 1318, App |
| 8 | 16 | 2.2 | 560 | US-09-270-767-16600 | Sequence 16600, A |
| 9 | 14 | 1.9 | 503 | US-09-270-767-27262 | Sequence 27262, A |
| 10 | 14 | 1.9 | 616 | US-09-328-111-332 | Sequence 332, App |
| 11 | 13 | 1.8 | 302 | US-09-313-294A-6869 | Sequence 6869, App |
| 12 | 12 | 1.7 | 265 | US-09-313-294A-910 | Sequence 910, App |

| | | | | | | |
|----|----|-----|------|---|---------------------|--------------------|
| 13 | 12 | 1.7 | 269 | 4 | US-09-313-294A-906 | Sequence 906, App |
| 14 | 12 | 1.7 | 292 | 4 | US-09-313-294A-4442 | Sequence 4442, App |
| 15 | 12 | 1.7 | 299 | 4 | US-09-313-294A-1303 | Sequence 1303, App |
| 16 | 12 | 1.7 | 307 | 4 | US-09-702-705-375 | Sequence 375, App |
| 17 | 12 | 1.7 | 307 | 4 | US-09-702-705-1271 | Sequence 1271, App |
| 18 | 12 | 1.7 | 307 | 4 | US-09-736-457-375 | Sequence 375, App |
| 19 | 12 | 1.7 | 307 | 4 | US-09-736-457-1271 | Sequence 1271, App |
| 20 | 12 | 1.7 | 307 | 4 | US-09-614-124B-375 | Sequence 375, App |
| 21 | 12 | 1.7 | 307 | 4 | US-09-614-124B-1271 | Sequence 1271, App |
| 22 | 12 | 1.7 | 307 | 4 | US-09-671-325-375 | Sequence 375, App |
| 23 | 12 | 1.7 | 307 | 4 | US-09-671-325-1271 | Sequence 1271, App |
| 24 | 12 | 1.7 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 25 | 12 | 1.7 | 307 | 4 | US-09-589-184-1271 | Sequence 1271, App |
| 26 | 12 | 1.7 | 307 | 4 | US-09-658-824-375 | Sequence 375, App |
| 27 | 12 | 1.7 | 487 | 3 | US-09-658-824-1271 | Sequence 1271, App |
| 28 | 12 | 1.7 | 487 | 3 | US-09-385-982-213 | Sequence 213, App |
| 29 | 12 | 1.7 | 487 | 4 | US-09-702-705-1655 | Sequence 1655, App |
| 30 | 12 | 1.7 | 487 | 4 | US-09-736-457-1655 | Sequence 1655, App |
| 31 | 12 | 1.7 | 487 | 4 | US-09-614-124B-1655 | Sequence 1655, App |
| 32 | 12 | 1.7 | 487 | 4 | US-09-671-325-1655 | Sequence 1655, App |
| 33 | 12 | 1.7 | 507 | 4 | US-09-658-824-1655 | Sequence 1655, App |
| 34 | 12 | 1.7 | 914 | 4 | US-09-513-999C-1965 | Sequence 1965, App |
| 35 | 12 | 1.7 | 1660 | 4 | US-09-270-767-12856 | Sequence 12856, A |
| 36 | 12 | 1.7 | 1660 | 4 | US-09-270-767-158 | Sequence 158, App |
| 37 | 12 | 1.7 | 1660 | 4 | US-09-270-767-15440 | Sequence 15440, A |
| 38 | 12 | 1.7 | 1660 | 4 | US-09-270-767-15440 | Sequence 15440, A |
| 39 | 11 | 1.5 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| 40 | 11 | 1.5 | 307 | 4 | US-09-736-457-588 | Sequence 588, App |
| 41 | 11 | 1.5 | 307 | 4 | US-09-614-124B-588 | Sequence 588, App |
| 42 | 11 | 1.5 | 307 | 4 | US-09-671-325-588 | Sequence 588, App |
| 43 | 11 | 1.5 | 307 | 4 | US-09-589-184-588 | Sequence 588, App |
| 44 | 11 | 1.5 | 307 | 4 | US-09-658-824-588 | Sequence 588, App |
| 45 | 10 | 1.4 | 109 | 4 | US-09-270-767-28703 | Sequence 28703, A |

ALIGNMENTS

RESULT 1

US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasuliojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganes M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614, 221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

| Alignment Scores: | Pred. No.: | Score: | Length: |
|------------------------|------------|---------------|---------|
| Best Local Similarity: | 0 | 631.00 | 2157 |
| Query Match: | 99.72% | Conservative: | 717 |
| DB: | 87.88% | Mismatches: | 0 |
| | 4 | Indels: | 1 |
| | | Gaps: | 2 |

US-10-028-384-6 (1-718) x US-09-614-221A-318 (1-2157)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleuLysLeuValIle 20
Db 1 ATGGATCCGACCGGTGCTGTGTTTGTCTGTGTTTTCAGACCATCTCAGCTCGTCATC 60

21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleIysPheGlu 40
61 TTCTGGTGGGCAATTTTGGGGCTGCCATATATCATCAGTTTGTTCAGTCATCAAAATTCAG 120
41 SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrIysIysTyrLeuValAsn 60
121 TCTATATTCATGATGATTCAGCCCTGGTTCATATATAGGGCTACCAAAATATCTGCTCAAC 180
61 AsnSerPheTyrLysPheLeuAsnTTPheAspAspArgTTPyrProLeuGlyArg 80
181 AATTCGTTTTCAGATTTTGAACCTGGTTTCAGACCGTACCTGGTACCCCTTCGGAAGG 240
81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
241 GTTACTGGGGGACTTATATCTCTGGTTTGTATGACGACTAGTGGCTTCATCTGGCAAGCC 300
101 LeuArgAsnTTPheLeuPheProIleAspIleArgAsnValCysValLeuPheAlaPro 120
301 CTGGCAACATGGTTGGGCTTCCCATTCACATCAGAAACGTTTGTGTGTATTTCTGGCCA 360
121 LeuPheSerGlyValThrAlaThrAlaThrTyrGluPheThrIysGluIleIysAspAla 140
361 CTATTTCTGGGGTCCACCGCTGGCGGACTTACGATTTACGAAAGAGATTAAGATGCC 420
141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
421 ACGGCTGGGCTTTTGGCTGGCTTTTATAGCCATTTGCCCGGTTATATATCTAGATCA 480
161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
481 GTGGCGGGTCTACGATAATAGGGCCATTGCCATTACACTATTAATGCTCACTTTCATG 540
181 PheThrIleLysAlaGlnIysThrGlySerIleMetHisAlaThrCysAlaAlaIlePhe 200
541 TTTTGGATTAAAGGCCAAAGACCTGGCTCTATCATGCAAGCAACGTTGTCAGCTTTATTC 600
201 TyrPheTyrMetValSerAlaThrPheGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
601 TACTTCTACATGGTGTGGCTGGGTGGGATACGTGTTTCATCACCACCTGATGCCACTC 660
221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerIysLeuTyrSerAlaTyrThr 240
661 CATGCTTTTCTGCTATTTGATGGGCGAGATATTCGTCACCAACTGATTCCTGCCACACC 720
241 ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
721 ACTTGGTACGCTATTTGGAACTTTGTCATCCATGCAATCCCATTTCTGCTTCTCTACCT 780
261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
781 ATCAGTCTTACGACCACTGGCCGCAATTTGGTGTGTTTCTGTTTGTATTCAGATTTGCGCC 840
281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
841 TTCGGTGTACTTCGTGAAGGGCCAAATCAGCACAGCTAAAGTTTAAAGTCTCATGATGTT 900
301 SerLeuPheLeuIleLeuValLeuGlyValGlyLeuSerAlaLeuThrTyrMetGly 320
901 TCTCTGTTTGTATCTGGTCTGTTGGTGTGGTGGGACTTCTGCCCTTGACCTATATGGGG 960
321 LeuIleAlaProTTPheThrGlyArgPheTyrSerLeuTTPheAspThrAsnTyrAlaIysIle 340
961 TTGATTTGCCCTTGGACTGGTAGATTTTATCTGGTATGGGATACCAACTACGCAAGATC 1020
341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPheAlaPhePhe 360
1021 CACATTTCTATCATTCCTCTGGTTCCGAACATCAACCCGTTTCTGTTGGCCGCTTCTTC 1080
361 PheAspThrHisPheLeuIleThrPhePheProAlaGlyValPheLeuLeuPheIleAsp 380
1081 TTTGTATACCCACTTTTGTATCTGGCTATTTCCCGCGGTGTATTTCTCTATTTCTCTGAC 1140
381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuLysSerTyrPheAlaGly 400

1141 TTGAAAGACGACGACGCTTTTGTTCATCGCTTACTCCGTTCTGTGTCTGTACTTTGCGGT 1200
401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
1201 GTTATGTTAGATTTGATGTTGACCTTGGACACCACTCATCTGTGTGTCGCGCGGTGCGA 1260
421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
1261 TTGTCCAGATATTTGACATCTACCTGGATTTCAAGCAAGTACCGCAAAATACGCCATC 1320
441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
1321 AAACCTGGCGCACTACTGCCAAATTTGATTTTCCGATCATTCATCTTTATTGTTAT 1380
461 LeuPheValPheHisSerThrTTPyrValThrArgThrAlaTyrSerSerProSerValVal 480
1381 CTTTTCGTTCTTCATCTTACTTTGGGTAAACAAGAACTGCATACTCTTCTCCTCTGTGT 1440
481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
1441 TTGCCATCACAAACCCAGATGGTAAATTTGGCGTTGATCGACGACTTCAGGGAAGCGTAC 1500
501 TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPheAspTyrGly 520
1501 TATTTGTTTAAAGATGAATCTCTGATGAGACAGTAAAGTTTGCAGCGTGGTGGGATTACGGT 1560
521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTTPheAsn 540
1561 TACCAAAATTTGGTGGCATGGCAGACAGAACCACTTTTAGTCGATAACAACACGTTGGAACAAT 1620
541 ThrHisIleAlaIleValIleValIleValIleValIleValIleValIleValIleVal 560
1621 ACTCATCTCGCATCTGTTGGTAAAGCCATGGCTTCCCTCAAGAGAAATCTTACGAAT 1680
561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
1681 CTAAAGAGCATGATGATGATTTGTTGTCATCTTGTGTCATCTTGTGTCATCTTGTGTTGT 1740
581 GlyAspAspIleAsnLysPheIleTTPheMetIleArgIleSerGluGlyIleTTPheProGlu 600
1741 GGTGATGATCATCAAAATTTCTTGTGATGATCAGAAATTCAGAGGGAATCTGGCCAGAA 1800
601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAl 620
1801 GAGATAAAGAGCGTGA-TTTCTATATCCGACAGAGGAGAAATACAGAGTAGATGCAAGGGC 1859
620 aSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLe 640
1860 TTCTGAGACCATGAGGAATCTGCTACTTTACAGATGTCCTACAGATTTCCCAAAATTTCCCAAT 1919
640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
1920 ATTCAATGTGGCCAAAGCACTGACAGAGTGGGTCAACAATGATCACCAATTTCCCAAT 1979
660 lProProLeuAspTyrPheAspGluValPheThrSerGluAsnTTPheMetValArgIleTyr 680
1980 CCCACCATTAGACTTCTGACGAGATTTTACTTCCGAAAACTGGATGGTAGAATA 2039
680 rGlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
2040 TCAATTTGAAGAGGATGATGCCAAGGTAGAACTTTGAGGAGCGTTGGTGAGTTAACCCAG 2099
700 gSerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
2100 GTCTTCTACGAAACACAGAGGTCCATAAAGAGACCTGAATTAGGCTTGAGAGTC 2154

RESULT 2

US-09-248-796A-3089

Sequence 3089, Application US/09248796A

Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 3089

; LENGTH: 867

; TYPE: DNA

; ORGANISM: Candida albicans

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (25)

; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown

US-09-248-796A-3089

Alignment Scores:
Pred. No.: 1,61e-21 Length: 867
Score: 30.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4,18% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-248-796A-3089 (1-867)

QY 515 AlaTPTpAspTyrGlyTyrGlnleGlyMetAlaAspArgThrThrLeuValasp 534

Db 244 GCCTGGTGGATTATGGTTATCAATCGGGGTATGCTGATAGAACACACTGGTGTAT 303

QY 535 AsnAsnThrTrpAsnAsnThrHisleAla 544

Db 304 AACATACATGATGAATACACATATATGCC 333

RESULT 3

US-09-248-796A-3090

; Sequence 3090.. Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 3090

; LENGTH: 594

; TYPE: DNA

; ORGANISM: Candida albicans

US-09-248-796A-3090

Alignment Scores:
Pred. No.: 1,6e-17 Length: 594
Score: 26.00 Matches: 26
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3,62% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-248-796A-3090 (1-594)

QY 66 PheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArgValThrGlyThr 85

Db 241 TTTTGAATGGTTTATGATAGACCTGGTACCCATGGAGAGTCACTGGTGTACT 300

QY 86 LeuTyrProGlyLeuMet 91
Db 301 TTATATCCCGTTTAAAG 318

RESULT 4

US-09-270-767-12331

; Sequence 12331, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 12331

; LENGTH: 900

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-12331

Alignment Scores:
Pred. No.: 4,13e-11 Length: 900
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2,79% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-12331 (1-900)

QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 532 TTCATCGCCATCGTCTGCTACATCAGTAGTCTGCTGATCGTACGATACGAG 591

RESULT 5

US-09-270-767-11648

; Sequence 11648, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 11648

; LENGTH: 1386

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-11648

Alignment Scores:
Pred. No.: 8,22e-08 Length: 1386
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2,37% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-11648 (1-1386)

QY 152 ILeValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168

Db 876 ATCGTCCCGGGTATATCTCTCGATCGTGGCGGATCGTACGACATGA 926

RESULT 6

US-09-513-999C-21090

; Sequence 21090, Application US/09513999C

; Patent No. 6783961

```
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; PATENT NO. 6783961
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 21090
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-21090

Alignment Scores:
Pred. No.: 1,7e-07 Length: 245
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.23% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-513-999C-21090 (1-245)
QY 149 PhelealalleValProGlyTyrIleSerArgSerValalaclySer 164
DB 60 TTTATTGCTATTGTACCAAGGCTACATATCTCGGTGATCGGTGGATCC-107

RESULT 7
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Alignment Scores:
Pred. No.: 3,76e-07 Length: 560
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.23% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-1318 (1-560)
QY 397 TyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 412
DB 402 TACTTCGCTGGTGTGATGGTGGTTTGAATGTTGACCTCAGCCGGTG 355

RESULT 8
US-09-270-767-16600/c
; Sequence 16600, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16600
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16600

Alignment Scores:
Pred. No.: 3,76e-07 Length: 560
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.23% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-16600 (1-560)
QY 397 TyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 412
DB 402 TACTTCGCTGGTGTGATGGTGGTTTGAATGTTGACCTCAGCCGGTG 355

RESULT 9
US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:
Pred. No.: 4,07e-05 Length: 503
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-270-767-27262 (1-503)
QY 155 GlyTyrIleSerArgSerValAlaGlySerTyrAsp&nglu 168
DB 2 GGGTATATCTCTCGATCCGTGGCGGATCTGTACGACATGAA 43

RESULT 10
US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endegre, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
```

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/068,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(616)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

Alignment Scores:
Pred. No.: 4.94e-05 Length: 616
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-6 (1-718) x US-09-328-111-332 (1-616)

Qy 201 TyPheTyrMetValSerAlaTrpGlyTyrValPheile 214
Db 149 TATTCTATATGCTCTCTGTTGGGTGGTTATGTTATTC 190

RESULT 11

US-09-313-294A-6869
; Sequence 6869, Application US/09313294A
; Patent No. 6476212

; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6869
; LENGTH: 302
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700352557H1/
US-09-313-294A-6869

Alignment Scores:
Pred. No.: 0.000272 Length: 302
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.81% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-6869 (1-302)

Qy 341 HisIleProfilLeaSerValSerGluHisGlnPro 353
Db 11 CATATACCAATCATGCATCTGTTTCTGAGCATGACCA 49

RESULT 12

US-09-313-294A-910
; Sequence 910, Application US/09313294A

; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 910
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700550053H1
US-09-313-294A-910

Alignment Scores:
Pred. No.: 0.00263 Length: 265
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.67% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-910 (1-265)

Qy 70 PheAspArgThrTrpTyrProLeuGlyArgVal 81
Db 3 TTTGATGATAGGACATGTACCCCTTGCCGTGTG 38

RESULT 13

US-09-313-294A-906
; Sequence 906, Application US/09313294A
; Patent No. 6476212

; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 906
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 700550043H1
US-09-313-294A-906

Alignment Scores:
Pred. No.: 0.00267 Length: 269
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.67% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-906 (1-269)

Qy 70 PheAspArgThrTrpTyrProLeuGlyArgVal 81
Db 3 TTTGATGATAGGACATGTACCCCTTGCCGTGTG 38

RESULT 14

US-09-313-294A-4442

us-10-028-384-6.Oligo.rn1

Thu Dec 16 16:25:08 2004

Db 256 ATCCCTATATTGCTAGTGTGAGCGACACCAACCA 291
 Search completed: December 14, 2004, 23:43:44
 Job time : 148.895 secs

; Sequence 4442, Application US/09313294A
 ; Patent No. 6476212
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath V.
 ; APPLICANT: Ito, Laura Y.
 ; APPLICANT: Sherman, Bradley K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
 ; FILE REFERENCE: PL-0017 US
 ; CURRENT APPLICATION NUMBER: US/09/313,294A
 ; CURRENT FILING DATE: 1999-05-14
 ; NUMBER OF SEQ ID NOS: 7600
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 4442
 ; LENGTH: 292
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6476212 700348434H1
 ; NAME/KEY: unsure
 ; LOCATION: 24
 ; OTHER INFORMATION: a, t, c, g, or other
 US-09-313-294A-4442

Alignment Scores:
 Pred. No.: 0.00289 Length: 292
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.67% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-4442 (1-292)

QY 533 ValAspAsnAsnThrTrpAsnAsnThrHisIleAla 544
 DB 150 GTTGATAACATACATGGTGAATACACACACATAGCT 185

RESULT 15

US-09-313-294A-1303
 ; Sequence 1303, Application US/09313294A
 ; Patent No. 6476212
 ; GENERAL INFORMATION:
 ; APPLICANT: Lalgudi, Raghunath V.
 ; APPLICANT: Ito, Laura Y.
 ; APPLICANT: Sherman, Bradley K.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
 ; FILE REFERENCE: PL-0017 US
 ; CURRENT APPLICATION NUMBER: US/09/313,294A
 ; CURRENT FILING DATE: 1999-05-14
 ; NUMBER OF SEQ ID NOS: 7600
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 1303
 ; LENGTH: 299
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Incyte ID No. 6476212 700550709H1
 US-09-313-294A-1303

Alignment Scores:
 Pred. No.: 0.00295 Length: 299
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.67% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x US-09-313-294A-1303 (1-299)

QY 342 IleProIleIleAlaSerValSerGluHisIlePro 353
 |||||||||||||||||||||||||||||||||||||||

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 : Search time 903.972 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 718

Sequence: 1 MGSDRSCLSVFQTLKLVI.....TRSTKTRRSIKPELGLRV 718

Scoring table:

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n:model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131648_6892/app_query.fasta_1.3740
-DB=Published Applications NA -OFT=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384.ecgn_1_12124@runat_14122004_131648_6892
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 718 | 100.0 | 2733 | 15 | US-10-028-384-5 |
| 2 | 631 | 87.9 | 2157 | 9 | US-09-804-368-387 |
| 3 | 631 | 87.9 | 2157 | 18 | US-10-793-639-318 |
| 4 | 32 | 4.5 | 2256 | 15 | US-10-032-585-6323 |
| 5 | 21 | 2.9 | 1848 | 15 | US-10-128-714-2139 |
| 6 | 21 | 2.9 | 1969 | 15 | US-10-128-714-1139 |
| 7 | 21 | 2.9 | 2232 | 15 | US-10-128-714-7139 |
| 8 | 21 | 2.9 | 2466 | 16 | US-10-320-797-2305 |
| 9 | 21 | 2.9 | 2603 | 15 | US-10-128-714-6139 |
| 10 | 21 | 2.9 | 2882 | 16 | US-10-320-797-1305 |
| 11 | 21 | 2.9 | 3969 | 15 | US-10-128-714-1319 |
| 12 | 21 | 2.9 | 4603 | 15 | US-10-128-714-5139 |
| 13 | 21 | 2.9 | 4738 | 16 | US-10-320-797-305 |
| 14 | 20 | 2.8 | 2417 | 15 | US-10-028-384-7 |
| 15 | 19 | 2.6 | 2681 | 16 | US-10-425-114-14408 |
| 16 | 19 | 2.6 | 2839 | 18 | US-10-425-115-150745 |
| 17 | 17 | 2.4 | 387 | 16 | US-10-276-774-173 |
| 18 | 17 | 2.4 | 393 | 16 | US-10-276-774-323 |
| 19 | 17 | 2.4 | 558 | 14 | US-10-052-283-433 |
| 20 | 17 | 2.4 | 1209 | 15 | US-10-106-698-330 |
| 21 | 17 | 2.4 | 1209 | 16 | US-10-264-237-412 |
| 22 | 17 | 2.4 | 1543 | 10 | US-09-974-879-133 |
| 23 | 17 | 2.4 | 1543 | 10 | US-09-305-736-133 |
| 24 | 17 | 2.4 | 1543 | 10 | US-09-818-683-133 |
| 25 | 17 | 2.4 | 1543 | 11 | US-09-818-683-133 |
| 26 | 17 | 2.4 | 1543 | 16 | US-10-621-401-133 |
| 27 | 17 | 2.4 | 2481 | 15 | US-10-028-384-1 |
| 28 | 17 | 2.4 | 2710 | 15 | US-10-028-384-3 |
| 29 | 16 | 2.2 | 154 | 10 | US-09-991-936-450 |
| 30 | 16 | 2.2 | 474 | 14 | US-10-052-283-454 |
| 31 | 16 | 2.2 | 485 | 10 | US-09-918-995-11283 |
| 32 | 16 | 2.2 | 500 | 9 | US-09-998-598-1643 |
| 33 | 16 | 2.2 | 1114 | 16 | US-10-296-115-629 |
| 34 | 16 | 2.2 | 2472 | 15 | US-10-171-581-112 |
| 35 | 16 | 2.2 | 2472 | 15 | US-10-028-384-11 |
| 36 | 16 | 2.2 | 2472 | 15 | US-10-172-118-742 |
| 37 | 16 | 2.2 | 2472 | 16 | US-10-342-887-742 |
| 38 | 16 | 2.2 | 3046 | 18 | US-10-417-375-95 |
| 39 | 16 | 2.2 | 3093 | 18 | US-10-417-375-92 |
| 40 | 16 | 2.2 | 3094 | 15 | US-10-028-384-9 |
| 41 | 16 | 2.2 | 5404 | 18 | US-10-417-375-99 |
| 42 | 16 | 2.2 | 5827 | 18 | US-10-417-375-97 |
| 43 | 16 | 2.2 | 54810 | 18 | US-10-417-375-91 |
| 44 | 16 | 2.2 | 103375 | 18 | US-10-417-375-94 |
| 45 | 15 | 2.1 | 528 | 18 | US-10-653-047-424 |

ALIGNMENTS

RESULT 1
US-10-028-384-5
; Sequence 5, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Kevin
; APPLICANT: MCERIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D28952
; DATABASE ENTRY DATE: 1999-02-07

; RELEVANT RESIDUES: (1) :.(2733)

US-10-028-384-5

Alignment Scores:

Pred. No.:

TEST NO.: 718-00
Score:

Score: 718.00
Percent similarity: 100.00%

Percent Similarity: 100.000
Best Local Similarity: 100.000

Best Local Similarity: 100.000

Query Match: 100.00%

DB: 15

US-10-028-384-6 (1-718) x US-10-028-384-5 (1-2733)

[illegible]

| | | | | | | | | | | | |
|----|------|--------|----------|--------|----------|--------|--------|---------|--------|----------|------|
| Db | 2199 | CCACCA | TAGACTAC | TCGACG | AGTTTTAC | CTTCG | AAAC | TGGATGG | TTAGAA | TATAT | 2258 |
| Qy | 681 | GlnLeu | LysAsp | AspAla | GlnGly | ArgThr | LeuArg | AspVal | GlyGlu | LeuThr | Arg |
| Db | 2259 | CAATTG | AGAAGA | GATGAT | GCCAAG | CTAGAA | CTTTG | GAGGAC | CTTGGT | GAGTTAAC | CGG |
| Qy | 701 | SerSer | ThrLys | ThrArg | ArgSer | LeuLys | ArgPro | GluLeu | ClyLeu | ArgVal | 718 |
| Db | 2319 | TCTTCT | TACGAA | CAACCA | GAGAGG | TCCTAA | AGAGAC | CTGAAT | TAGGCT | TGAGATC | 2372 |

RESULT 2

```

US-09-801-368-397
; Sequence 387, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 108272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 387
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-387

```

| Alignment Scores: | |
|---------------------|--------|
| Aligned. No.: | 0 |
| Score: | 631.00 |
| Percent Similarity: | 99.72% |
| Local Similarity: | 99.72% |
| Query Match: | 87.88% |
| DB: | 9 |
| Gaps: | 0 |
| Indels: | 2 |
| Mismatches: | 1 |
| Conservative: | 0 |
| Matches: | 2157 |
| Length: | 2157 |

US-10-028-384-6 (1-718) x US-09-801-368-387 (1-2157)

| | | |
|-----|--|-----|
| 1 | MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle | 20 |
| 1 | ATGGGATCCGACCGGTGGTGTGTGTGTGTGTTCAGACCATCTCAAGCTCGTCATC | 60 |
| 21 | PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu | 40 |
| 61 | TTCGTGGCGATTTTGGGGTGCATATCATCACTGTTGTTTGCAGTCATCAAAATTCAG | 120 |
| 41 | SerIleIleHisGluPheAspProTppPheAsnTyrArgAlaThrLysTyrLeuValAsn | 60 |
| 121 | TCTATTATCCATGAATTCGACCCCTGGTTCATTATAGGGCTACCAATATCTGTCAC | 180 |
| 61 | AsnSerPheTyrLysPheLeuAsnTppPheAspAspArgThrTrpTyrProLeuGlyArg | 80 |
| 181 | AAATCGTTTACAAGTTTTGAACGTGGTTTGACGCCGACTCGTGACCCCTCGGAAGG | 240 |
| 81 | ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAla | 100 |

QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerProSerValVal 480
 Db 1381 CTTTTCGCTTCCTACTTCTGGTAAACAGAACTGCATCTCTCTCTCTCTCTCT 1440
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuLeuLeuLeuLeuLeu 500
 Db 1441 TTGCCATCAACACCCAGATGTAATTTGGCTTCATCGACGACTTCAGGGAACGATC 1500
 QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaAlaAlaAla 520
 Db 1501 TATTGGTTAAGATGAATCTCTGATGAGGACAGTAAGTTTGCAGCTGGTGGATACGGT 1560
 QY 521 TyrGlnThrGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsn 540
 Db 1561 TACCAGATTTGGTCATGGCAGACAGACACATTTAGTCGATAACACACGTTGGAAAT 1620
 QY 541 ThrHisLeuAlaLeuValGlyLysAlaMetAlaSerProGluGluLysSerTyrGlu 560
 Db 1621 ACTCACATGCCATCGTTGTTAAGCCATGGCTTCCCTGAAGAGAAATCTTACGAAT 1680
 QY 561 LeuLysGluHisAspValAspTyrValLeuValLeuValLeuValLeuValLeu 580
 Db 1681 CTAAAGAGAGATGATGTCGATTAATGCTTGGTTCATCTTTGGTGGTCTAATTTGGGTTGGT 1740
 QY 581 GlyAspAspLeuAsnLysPheLeuTrpMetIleArgIleSerGluGlyLysIleTrpPro 600
 Db 1741 GGTGATGACATCAACAAATTTCTGTGATGATCAGAAATTTAGCGAGGAATCTGGCCAGAA 1800
 QY 601 GluLysGluArgTyr-PheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAl 620
 Db 1801 GAGATAAAGAGCGTGA-TTTCATACCCGAGGAGGAGAAATACAGAGTAGATGCAAGGCG 1859
 QY 620 aSerGluThrMetArgAsnSerLeuLysMetSerTyrLysAspPheProGlnLe 640
 Db 1860 TTCTGAGACCATGAGGAATCGCTACTTTACAGATGCTCTACAAAGATTTCCCAAT 1919
 QY 640 uPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVa 660
 Db 1920 ATTCAATGGTGGCCAGCCACTGACAGAGTGGCTCAACAAATGATCACCATTAGACGT 1979
 QY 660 lProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleT 680
 Db 1980 CCCACCATTAGACTACTTCGACCAAGTTTACTTCCGAAACTGGATGGATTAGATATA 2039
 QY 680 rGlnLeuLysLysAspAspAlaGlnArgThrLeuArgAspValGlyLeuThrArg 700
 Db 2040 TCAATTGAAGAGGATGATGCCCAAGGTAGAACTTTGAGGGACGTTGGTGGATTAAACAG 2099
 QY 700 gSerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2100 GTCTTCTACGAAACAGAGGTCCTAAGAGACCTGAATTAGGCTTGAGATC 2154

RESULT 3

US-10-793-639-318
 ; Sequence 318, Application US/10793639
 ; Publication No. US20040199940A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kartunandaa, Balasulajini
 ; APPLICANT: Yu, Jaehyuk
 ; APPLICANT: Kishore, Ganesh M.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
 ; WITH STEROL SYNTHESIS AND METABOLISM
 ; FILE REFERENCE: 16516.075
 ; CURRENT APPLICATION NUMBER: US/10/793,639
 ; PRIOR FILING DATE: 2004-03-05
 ; PRIOR APPLICATION NUMBER: US/09/614,221A
 ; PRIOR FILING DATE: 2000-07-11
 ; PRIOR APPLICATION NUMBER: US 60/142,981
 ; NUMBER OF SEQ ID NOS: 626
 ; SEQ ID NO 318
 ; LENGTH: 2157

TYPE: DNA
 ORGANISM: Saccharomyces cerevisiae
 US-10-793-639-318

Alignment Scores:

Pred. No.: 0 Length: 2157
 Score: 631.00 Matches: 717
 Percent Similarity: 99.72% Conservative: 1
 Best Local Similarity: 99.72% Mismatches: 2
 Query Match: 87.88% Indels: 2
 DB: 18 Gaps: 0

US-10-028-384-6 (1-718) x US-10-793-639-318 (1-2157)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValLe 20
 Db 1 ATGGGATCGACCGGTCGT 60
 QY 21 PheValAlaIlePheGlyValAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTCGTGGCGATTTTGGGGCTGCCATATCATCAGTTTGTTCGATCATCAATTTGAG 120
 QY 41 SerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 Db 121 TCTATTATCCATGAATTCGACCCCTGTGTCAATATTATAGGGCTACCAATATATCTGTC 180
 QY 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspAspArgThrTyrTyrProLeuGlyArg 80
 Db 181 AATTGGTTTACAGTTTGTGAACCTGTTTGCAGCCGCTGCTGTTACCCCTCGGAGG 240
 QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrPheAla 100
 Db 241 GTTACTGGAGGACATTTATATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
 QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGGCCCACTGGTTGGGCTTCCCATGTGACATCAGAAACGTTTGTGTGTGTGTGTGTGT 360
 QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 Db 361 CTATTTCCTGGGCTCACCGCTCGGCGACATTACGAATTTACGAAGAGATTAAGATGCC 420
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 421 AGCGTGGGCTTTGGCTGT 480
 QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGCGGGGCTCTACGATAATGAGGCCATTGCGCATTTACACTATTATTAATGGTCACCT 540
 QY 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAGGCCCAAAAGACTGGCTCTATCATGCGCAAGCTGTGGAGCTTTATTTC 600
 QY 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTGTGGCTTGGGCTGGATGATGTTTATCATCCACCACTTGTATCCCA 660
 QY 221 HisValPheLeuLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 661 CATGCTTTTGT 720
 QY 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGTGATCGTATTGGAACATGTTGTCATCCATGAGATCCCATTTGTGTGTGTGTGTGT 780
 QY 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGTCTACGACCAACATGCGCATTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 840
 QY 281 PheGlyAspPheValGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TTCGTGACTTCTGTGAAGGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGTGTGT 900

Thu Dec 16 16:25:08 2004

; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 2139
 ; LENGTH: 1848
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-2139

Alignment Scores:
 Pred. No.: 2,18e-11 Length: 1848
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-2139 (1-1848)

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 448 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGAACCTGATCCCTGCACGTT 507
 QY 223 Phe 223
 DB 508 TTT 510

RESULT 6

US-10-128-714-1139
 ; Sequence 1139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengdi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 1139
 ; LENGTH: 1969
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-1139
 Alignment Scores:
 Pred. No.: 2.32e-11 Length: 1969
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0
 US-10-028-384-6 (1-718) x US-10-128-714-1139 (1-1969)

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 514 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGAACCTGATCCCTGCACGTT 573
 QY 223 Phe 223
 DB 574 TTT 576

RESULT 7

US-10-128-714-7139
 ; Sequence 7139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengdi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 7139
 ; LENGTH: 2232
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-7139
 Alignment Scores:
 Pred. No.: 2.61e-11 Length: 2232
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0
 US-10-028-384-6 (1-718) x US-10-128-714-7139 (1-2232)
 QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 DB 625 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGAACCTGATCCCTGCACGTT 684
 QY 223 Phe 223
 DB 685 TTT 687

RESULT 8

US-10-320-797-2305
 ; Sequence 2305, Application US/10320797
 ; Publication No. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eroshkin, Alexey M.
 ; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
 ; TITLE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; CURRENT FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2305
; LENGTH: 2466
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
US-10-320-797-2305

Alignment Scores:
Pred. No.: 2,87e-11 Length: 2466
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 16 Gaps: 0

US-10-028-384-6 (1-718) x US-10-320-797-2305 (1-2466)

QY 152 ileValProGlyTyrIleSerArgSerValalacGlySerTyrAspAsnGluAla 171
Db 637 ATTGTACCGGATACATCTCGATCTGTCGCGGTTCTTATGACACGAAGCAATTGCC 696

QY 172 ile 172
Db 697 ATC 699

RESULT 9

US-10-128-714-6139
; Sequence 6139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES IN ASPERGILLUS FUMIGATUS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6139
; LENGTH: 2603
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-6139

Alignment Scores:
Pred. No.: 3,02e-11 Length: 2603
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-6139 (1-2603)

QY 203 TyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeulleProLeuHisVal 222

Db 877 TACATGGTGTGGCATGGGTGGGTATGTTTCATTACGAACCTGATCCCCCTGCACGT 936
QY 223 Phe 223
Db 937 TTT 939

RESULT 10

US-10-320-797-1305
; Sequence 1305, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1305
; LENGTH: 2882
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
US-10-320-797-1305

Alignment Scores:

Pred. No.: 3,33e-11 Length: 2882
Score: 21.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.92% Indels: 0
DB: 16 Gaps: 0

US-10-028-384-6 (1-718) x US-10-320-797-1305 (1-2882)

QY 152 ileValProGlyTyrIleSerArgSerValalacGlySerTyrAspAsnGluAla 171
Db 796 ATTGTACCGGATACATCTCTCGATCTGTCGCGGTTCTTATGACACGAAGCAATTGCC 855

QY 172 ile 172
Db 856 ATC 858

RESULT 11

US-10-128-714-139
; Sequence 139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES IN ASPERGILLUS FUMIGATUS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 139
 ; LENGTH: 3969
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-139

Alignment Scores:
 Pred. No.: 4,51e-11 Length: 3969
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-139 (1-3969)

Qy 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 1514 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGACCTGATCCCGCTGCACGTT 1573

Qy 223 Phe 223
 Db 1574 TTT 1576

RESULT 12

US-10-128-714-5139
 ; Sequence 5139, Application US/10128714
 ; Publication No. US20030119013A1

; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wenqi
 ; APPLICANT: Tishkoff, Daniel

; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Eroshkin, Alexey M

; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23
 ; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27

; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5139
 ; LENGTH: 4603

; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus

US-10-128-714-5139

Alignment Scores:
 Pred. No.: 5.19e-11 Length: 4603
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 15 Gaps: 0

US-10-028-384-6 (1-718) x US-10-128-714-5139 (1-4603)

Qy 203 TyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisVal 222
 Db 1877 TACATGGTGTGGCATGGGTGGGTATGCTTCATTACGACCTGATCCCGCTGCACGTT 1936

Qy 223 Phe 223
 Db 1937 TTT 1939

RESULT 13

US-10-320-797-305

; Sequence 305, Application US/10320797
 ; Publication No. US20040014955A1

; GENERAL INFORMATION:
 ; APPLICANT: Eroshkin, Alexey M.

; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND

; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797

; CURRENT FILING DATE: 2002-12-16
 ; PRIOR FILING DATE: 2001-12-17

; PRIOR APPLICATION NUMBER: 60/341,261
 ; NUMBER OF SEQ ID NOS: 3361

; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 305

; LENGTH: 4738
 ; TYPE: DNA

; ORGANISM: Cryptococcus neoformans
 US-10-320-797-305

Alignment Scores:

Pred. No.: 5.34e-11 Length: 4738
 Score: 21.00 Matches: 21
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 2.92% Indels: 0
 DB: 16 Gaps: 0

US-10-028-384-6 (1-718) x US-10-320-797-305 (1-4738)

Qy 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 Db 1652 ATTGTACCGGATACATCTCTCGATCTGTGCGCGTCTTATGACACGAAGCCATTGCC 1711

Qy 172 Ile 172
 Db 1712 ATC 1714

RESULT 14

US-10-028-384-7

; Sequence 7, Application US/10028384
 ; Publication No. US20030148285A1

; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.

; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCBRIDE, Kevin

; TITLE OF INVENTION: Mammalian SIMP, Gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74

; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7
 ; LENGTH: 2417

; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster

; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: AF132552

; DATABASE ENTRY DATE: 1999-04-27
 ; RELEVANT RESIDUES: (1)..(2417)

US-10-028-384-7

Alignment Scores:

Pred. No.: 3.18e-10 Length: 2417
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservativeness: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 14, 2004, 13:16:58 ; Search time 5558.03 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 718
Sequence: 1 MGSRSCLVSVFQTLKLV.....TRSTKRSIKRPELGLRV 718

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 32822875 seqs, 18219865908 residues

Word size: 1
Total number of hits satisfying chosen parameters: 65644297
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10028384/runat_14122004_131646_6807/app_query.fasta_1.3740
-DB=EST -QFMT=fastcap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384@cgn_1.13960 @runat_14122004_131646_6807 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -YGAPEXT=60 -Fgapop=6
-Fgapext=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 47 | 5.5 | 981 | 9 | CNS06Y0T |
| 2 | 41 | 5.7 | 869 | 9 | CNS06DY5 |
| 3 | 41 | 5.7 | 1088 | 9 | CNS06CLU |
| 4 | 35 | 4.9 | 463 | 8 | BZ301622 K01378.pl |
| 5 | 32 | 4.5 | 972 | 9 | CNS07A8F |
| 6 | 29 | 4.0 | 1074 | 9 | CNS06ZDA |
| 7 | 25 | 3.5 | 593 | 8 | BZ295215 |
| 8 | 24 | 3.3 | 950 | 9 | CNS06L1S |
| 9 | 24 | 3.3 | 950 | 9 | CNS06L1W |

| | | | | | |
|----|----|-----|-----|---|----------|
| 10 | 20 | 2.8 | 424 | 7 | CO337666 |
| 11 | 20 | 2.8 | 576 | 4 | EM250810 |
| 12 | 20 | 2.8 | 585 | 1 | AI133998 |
| 13 | 20 | 2.8 | 594 | 4 | BM638349 |
| 14 | 20 | 2.8 | 598 | 4 | BI170396 |
| 15 | 20 | 2.8 | 600 | 4 | BI163592 |
| 16 | 20 | 2.8 | 615 | 4 | BI364613 |
| 17 | 20 | 2.8 | 616 | 2 | BF504378 |
| 18 | 20 | 2.8 | 630 | 1 | AI295381 |
| 19 | 20 | 2.8 | 635 | 1 | AI135629 |
| 20 | 20 | 2.8 | 641 | 1 | AI257750 |
| 21 | 20 | 2.8 | 642 | 4 | BI173041 |
| 22 | 20 | 2.8 | 644 | 4 | BG636414 |
| 23 | 20 | 2.8 | 649 | 4 | BG641064 |
| 24 | 20 | 2.8 | 652 | 4 | BI171940 |
| 25 | 20 | 2.8 | 654 | 4 | BI374334 |
| 26 | 20 | 2.8 | 655 | 2 | BF496296 |
| 27 | 20 | 2.8 | 655 | 4 | BI484774 |
| 28 | 20 | 2.8 | 658 | 4 | BI374189 |
| 29 | 20 | 2.8 | 663 | 4 | BG641172 |
| 30 | 20 | 2.8 | 666 | 4 | BI357074 |
| 31 | 20 | 2.8 | 668 | 4 | BI27902 |
| 32 | 20 | 2.8 | 679 | 7 | CK657649 |
| 33 | 20 | 2.8 | 681 | 7 | CK659033 |
| 34 | 20 | 2.8 | 682 | 7 | CK659064 |
| 35 | 20 | 2.8 | 700 | 4 | BM595496 |
| 36 | 20 | 2.8 | 715 | 7 | CK657451 |
| 37 | 20 | 2.8 | 718 | 6 | CD779819 |
| 38 | 20 | 2.8 | 736 | 1 | AA949890 |
| 39 | 20 | 2.8 | 784 | 6 | CD783059 |
| 40 | 20 | 2.8 | 788 | 6 | CD781399 |
| 41 | 20 | 2.8 | 810 | 2 | BF502026 |
| 42 | 20 | 2.8 | 876 | 6 | CD794446 |
| 43 | 20 | 2.8 | 878 | 6 | CD788585 |
| 44 | 20 | 2.8 | 891 | 6 | CD778215 |
| 45 | 20 | 2.8 | 960 | 6 | CD793779 |

ALIGNMENTS

RESULT 1

CNS06Y0T

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

CNS06Y0T 981 bp DNA linear GSS 06-JUL-2001
T3 end of clone AY0AA005C11 of library AY0AA from strain CBS 6340
of Kluyveromyces thermotolerans, genomic survey sequence.

AL420435

AL420435.1 GI:12203620

GSS

Kluyveromyces thermotolerans

Kluyveromyces thermotolerans

Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

1 (bases 1 to 981)

Souciot, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bcn, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Portier, S.,

Saurin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M.,

Wincker, P. and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEMS Lett. 487 (1), 3-12 (2000)

20584711

11152876

2 (bases 1 to 981)

Malpertuy, A., Llorente, B., Blandin, G., Artiguenave, F., Wincker, P.

and Dujon, B.

Genomic exploration of the hemiascomycetous yeasts: 10.

Kluyveromyces thermotolerans

FEMS Lett. 487 (1), 61-65 (2000)

20584720

11152885

3 (bases 1 to 981)

| | |
|----------|------------|
| CO337666 | EN15033.5 |
| EM250810 | EST577344 |
| AI133998 | GHI1327.5 |
| BM638349 | I70006875 |
| BI170396 | REI11825.5 |
| BI163592 | REO3026.5 |
| BI364613 | REA9773.5 |
| BF504378 | AT05729.5 |
| AI295381 | LP08987.5 |
| AI135629 | GH13452.5 |
| AI257750 | LP06212.5 |
| BI173041 | REI5774.5 |
| BG636414 | SD14123.5 |
| BG641064 | SD12296.5 |
| BI171940 | REI13841.5 |
| BI374334 | REI1893.5 |
| BF496296 | AT10060.5 |
| BI484774 | REB7926.5 |
| BI374189 | REB1718.5 |
| BG641172 | SD12448.5 |
| BI357074 | RE43425.5 |
| BI27902 | RE25288.5 |
| CK657649 | LP24119.5 |
| CK659033 | LP16462.5 |
| CK659064 | LP16523.5 |
| BM595496 | I70006874 |
| CK657451 | LP23759.5 |
| CD779819 | EST651180 |
| AA949890 | LD29946.5 |
| CD783059 | EST654420 |
| CD781399 | EST652760 |
| BF502026 | AT17637.5 |
| CD794446 | EST665807 |
| CD788585 | EST658946 |
| CD778215 | EST649576 |
| CD793779 | EST665140 |


```

SOURCE      Zygosaccharomyces rouxii
ORGANISM    Zygosaccharomyces rouxii
REFERENCE   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS     Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.
1 (bases 1 to 1088)
Souciet, J.-L., Aglie, M., Artiguenave, F., Blandin, G.,
Botot-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrien, P., Lepingle, A., Llorente, B.,
Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
REFERENCE   2 (bases 1 to 1088)
AUTHORS     de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B.,
Wincker, P., Artiguenave, F. and Souciet, J.
Genomic exploration of the hemiascomycetous yeasts: 8.
Zygosaccharomyces rouxii
FEBS Lett. 487 (1), 52-55 (2000)
20584718
11152883
REFERENCE   3 (bases 1 to 1088)
AUTHORS     Genoscope.
Direct Submission
Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and farrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES    Location/Qualifiers
            source
            1..1088
            /organism="Zygosaccharomyces rouxii"
            /mol_type="genomic DNA"
            /strain="CBS 732"
            /db_xref="taxon:4956"
            /clone="ARO0A006C05"
            /clone_lib="AR0AA"
            /note="end : 17"
            misc_feature
            complement( <3..>1006)
            /note="similar to Saccharomyces cerevisiae ORF YGL022w [
            SRT3 ; Oligosaccharyl transferase subunit ]
            1 putative frameshift (8)"
            /evidence=not_experimental

ALIGNMENT Scores:
Pred. No.:      9,398-32      Length:      1088
Score:          41.00      Matches:      41
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    5.71%      Indels:      0
DB:             9      Gaps:      0

US-10-028-384-6 (1-718) x CNS06CLU (1-1088)

QY 151 AlalIeValPrCGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIle 170
Db 566 GCGATTGTGCCAGGTTCATTCTAGATCAGTTCAGGTCCTTACGATAACGAGCTATT 507
QY 171 AlalIeThrLeuLeuMetValThrPheMetPheTyrIleLysAlaGlnLysThrGlySer 190
Db 506 GCAATTACCCCTBTAAATGGTTTACTTTTCATGTTCTCGGATTAAAGCTCAAAAGACAGGTTCT 447

SOURCE      Zygosaccharomyces delphensis
ORGANISM    Zygosaccharomyces delphensis
REFERENCE   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS     Kluyveromycetales; Saccharomycetaceae; Kluyveromycetes.
1 (bases 1 to 463)
Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.
Evidence from comparative genomics for a complete sexual cycle in
the 'asexual' pathogenic yeast Candida glabrata
Genome Biol. 4 (2), R10 (2003)
22508158
12620120
COMMENT     Contact: Wong S
            Department of Genetics, Smurfit Institute
            Trinity College Dublin
            Dublin 2, Ireland
            Tel: 353 1 6082319
            Fax: 353 1 6798558
            Email: swong@tcd.ie
            Class: plasmid ends.
FEATURES    Location/Qualifiers
            source
            1..463
            /organism="Kluyveromyces delphensis"
            /mol_type="genomic DNA"
            /strain="CBS 2170"
            /db_xref="taxon:51657"
            /clone="KD1378"
            /clone_lib="Kluyveromyces delphensis Random Genomic
            Library"

ALIGNMENT Scores:
Pred. No.:      8,498-26      Length:      463
Score:          35.00      Matches:      35
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    4.87%      Indels:      0
DB:             8      Gaps:      0

US-10-028-384-6 (1-718) x BZ301622 (1-463)

QY 320 GlyLeuIleAlaProTyrThrGlyArgPheTyrSerLeuTyrPaspThrAsnTyrAlaLys 339
Db 160 GGTTCGATCGCGCCATCGACTGGAGAGATTTTACTCTTTGGGATACCAATTACGCTAAG 219
QY 340 IleHisIleProIleIleAlaSerValSerGluHisGluProVal 354
Db 220 ATTCAATCCCAATCAATGGTCTGTCTCTGAAACATCAACCTGTT 264

RESULT 5
CNS07A8F
LOCUS
DEFINITION T7 end of clone BC0AA002D03 of library BC0AA from strain CBS 767 of
            Debaryomyces hansenii, genomic survey sequence.
ACCESSION AL436261
VERSION   AL436261.1
KEYWORDS  GI:12219674
SOURCE    Debaryomyces hansenii (anamorph: Candida famata)
ORGANISM  Debaryomyces hansenii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

Saccharomycetales; Saccharomycetaceae; Debaryomyces.
1 (bases 1 to 972)
LOCUS      Saccharomycetaceae; Debaryomyces.
DEFINITION
ACCESSION  Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
VERSION    Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
AL422180   de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
KEYWORDS   Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
SOURCE     Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
ORGANISM   Wincker,P. and Weissenbach,J.
REFERENCE  1 (bases 1 to 972)
AUTHORS    Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE    20584711
PUBMED     11152876
REFERENCE  2 (bases 1 to 972)
AUTHORS    Lepingle,A., Casaregola,S., Neuveglise,C., Bon,E., Nguyen,H.,
            Artiguenave,F., Wincker,P. and Gallardin,C.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 14.
            Debaryomyces hansenii var. hansenii
JOURNAL    FEBS Lett. 487 (1), 82-86 (2000)
MEDLINE    20584724
PUBMED     11152889
REFERENCE  3 (bases 1 to 972)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT    This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES   Location/Qualifiers
            source          1..972
                        /organism="Debaryomyces hansenii"
                        /mol_type="genomic DNA"
                        /strain="CBS 767"
                        /variety="hansenii"
                        /db_xref="taxon:4959"
                        /clone="BC0AA002D03"
                        /clone_lib="BC02A"
                        /note="end : T7"
            misc_feature    <5..>970
                        /note="similar to Saccharomyces cerevisiae ORF YGL022w [
                        STT3 ; oligosaccharyl transferase subunit ]"
                        /evidence=not_experimental
ORIGIN
Alignment Scores:
Pred. No.:      2,59e-22      Length:      972
Score:          32.00        Matches:      32
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    4.46%        Indels:      0
DB:             9           Gaps:        0

US-10-028-384-6 (1-718) x CNS07A8F (1-972)

QY      322  ILeaLaPrCTpThrGlyArgGPhETySerLeuTipAspThrAsnTyAlaIyAlaIleHis 341
Db      143  ATTGCACCTTGGATGCGCGTTCTATCTTTATGGATCACTAACTATCCCAAGATTCAC 202
QY      342  ILeProIleLeaIaSerValSerGluHisGlnPro 353
Db      203  ATTCTATATATGTCATCGTTCCTTGTGAACATCAACCT 238

RESULT 6
CNS06ZDA

```

```

CNS06ZDA      1074 bp      DNA      linear      GSS 06-JUL-2001
T3 end of clone XAY0AA001F10 of library XAY0AA from strain CBS 6340
of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION  AL422180
VERSION    AL422180.1 GI:12205374
KEYWORDS   GSS.
SOURCE     Kluyveromyces thermotolerans
ORGANISM   Kluyveromyces thermotolerans
REFERENCE  1 (bases 1 to 1074)
AUTHORS    Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaiia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL    FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE    20584711
PUBMED     11152876
REFERENCE  2 (bases 1 to 1074)
AUTHORS    Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
            and Dujon,B.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 10.
            Kluyveromyces thermotolerans
JOURNAL    FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE    20584720
PUBMED     11152885
REFERENCE  3 (bases 1 to 1074)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (09-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT    This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES   Location/Qualifiers
            source          1..1074
                        /organism="Kluyveromyces thermotolerans"
                        /mol_type="genomic DNA"
                        /strain="CBS 6340"
                        /db_xref="taxon:4916"
                        /clone="XAY0AA001F10"
                        /clone_lib="XAY0AA"
                        /note="end : T3"
            misc_feature    4882..>1066
                        /note="similar to Saccharomyces cerevisiae ORF YGL022w [
                        STT3 ; oligosaccharyl transferase subunit ]"
                        /evidence=not_experimental
ORIGIN
Alignment Scores:
Pred. No.:      4,16e-19      Length:      1074
Score:          29.00        Matches:      29
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    4.04%        Indels:      0
DB:             9           Gaps:        0

US-10-028-384-6 (1-718) x CNS06ZDA (1-1074)

QY      63  PhETyLysePhLeuAsnTrpPheAspArgThrTrpTyProLeuGlyArgValThr 82
CNS06ZDA

```

```

Db      835 TTTTATAAGTCTCTTAATGTTGTCAGACAGACTTGGTATCCCTGGCAGAGTCACC 894
QY      83  GYGLVThrLeuTyPProGlyLeuMet 91
Db      895 GGTGGCAGCGTGTATCCCGGTGTATG 921

RESULT 7
BZ295215/c
LOCUS      593 bp DNA linear GSS 31-OCT-2002
DEFINITION Candida glabrata Random Genomic Library Candida glabrata
ACCESSION BZ295215
VERSION    BZ295215.1 GI:24436329
KEYWORDS   GSS.
SOURCE     Candida glabrata
ORGANISM   Candida glabrata
REFERENCE  1 (bases 1 to 593)
AUTHORS    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
TITLE      Candida glabrata
JOURNAL    Saccharomycetales; mitosporic Saccharomycetales; Candida.
MEDLINE    1 (bases 1 to 593)
PUBMED     12620120
AUTHORS     Contact: Wong S
TITLE       Department of Genetics, Smurfit Institute
JOURNAL     Trinity College Dublin
MEDLINE     Dublin 2, Ireland
PUBMED     Tel: 353 1 6082319
AUTHORS     Fax: 353 1 6798558
TITLE       Email: swong@tcd.ie
JOURNAL     Class: plasmid ends.
MEDLINE     Location/Qualifiers
PUBMED     1..593
AUTHORS     /organism="Candida glabrata"
TITLE       /mol_type="genomic DNA"
JOURNAL     /strain="CBS 138"
MEDLINE     /db_xref="taxon:5478"
PUBMED     /clone="CG1357"
COMMENT     /clone_lib="Candida glabrata Random Genomic Library"

FEATURES
source
1..593
Location/Qualifiers
/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS 138"
/db_xref="taxon:5478"
/clone="CG1357"
/clone_lib="Candida glabrata Random Genomic Library"

ORIGIN
Alignment Scores:
Pred. No.: 3..796-15 Length: 593
Score: 25.00 Matches: 25
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3..48% Indels: 0
DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x BZ295215 (1-593)

QY 527 A'AspArgThrThrLeuValAspAsnAsnThrTTPAsnAsnThrHisIleAlaIleVal 546
Db 481 GCAGACAGAACCACTTGTGTTGACAACTACCTCGGAACAATACACATATCGCAATGTC 422
QY 547 GYLYSAlaMetAla 551
Db 421 GGTAAAGCAATGGCT 407

RESULT 8
CNS06L1S
LOCUS      950 bp DNA linear GSS 17-JUN-2001
DEFINITION T3 end of clone AT0AA009H01 of library AT0AA from strain CBS 4311
ACCESSION AL403622
VERSION    AL403622.1 GI:12164087
KEYWORDS   GSS.
SOURCE     Saccharomycetes servazzii
ORGANISM   Saccharomycetes servazzii
REFERENCE  1 (bases 1 to 950)
AUTHORS    Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
TITLE      Genomic exploration of the hemiascomycetous yeasts: 1. A set of
JOURNAL    yeast species for molecular evolution studies
MEDLINE    FEBS Lett. 487 (1), 3-12 (2000)
PUBMED     20584711
AUTHORS     2 (bases 1 to 950)
TITLE      Casaregola, S., Lepingle, A., Bon, E., Neuveglise, C., Nguyen, H.,
JOURNAL    Artiguenave, P., Wincker, P. and Gaillardin, C.
MEDLINE    Genomic exploration of the hemiascomycetous yeasts: 7.
PUBMED     Saccharomycetes servazzii
AUTHORS     FEBS Lett. 487 (1), 47-51 (2000)
TITLE      20584717
JOURNAL    Genoscope.
MEDLINE    3 (bases 1 to 950)
PUBMED     Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
AUTHORS     2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
TITLE      sefref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
JOURNAL    This GSS is part of a random genomic sequencing program of thirteen
MEDLINE    yeast species: Saccharomycetes bayanus var. uvarum, Saccharomycetes
PUBMED     exiguus, Saccharomycetes servazzii, Zygosaccharomycetes rouxii,
AUTHORS     Saccharomycetes kluyveri, Kluyveromycetes thermotolerans, Kluyveromycetes
TITLE      lactis var. lactis, Kluyveromycetes marxianus var. marxianus, Pichia
JOURNAL    angusta, Debaryomycetes hansenii var. hansenii, Pichia sorbitophila,
MEDLINE    Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
PUBMED     5 kb were prepared and both extremities were sequenced. See
AUTHORS     keywords for description of this sequence and for the sequence of
TITLE      the other extremity of this insert.
JOURNAL    Location/Qualifiers
MEDLINE    1..950
PUBMED     /organism="Saccharomycetes servazzii"
AUTHORS     /mol_type="genomic DNA"
TITLE      /strain="CBS 4311"
JOURNAL    /db_xref="taxon:27293"
MEDLINE    /clone="AT0AA009H01"
PUBMED     /clone_lib="AT0AA"
AUTHORS     /notes="end : T3"
TITLE      <95..>500
JOURNAL    /notes="similar to Saccharomycetes cerevisiae ORF YGL022w {
MEDLINE    STT3 ; oligosaccharyl transferase subunit }
PUBMED     1 putative frameshift(s)"
AUTHORS     /evidence=not_experimental
TITLE      complement(<837..>948)
JOURNAL    /notes="similar to Saccharomycetes cerevisiae ORF YGL020c {
MEDLINE    hypothetical protein }
PUBMED     /evidence=not_experimental
COMMENT     /evidence=not_experimental

ORIGIN
Alignment Scores:
Pred. No.: 6..88e-14 Length: 950
Score: 24.00 Matches: 24
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3..34% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-6 (1-718) x CNS06L1S (1-950)

QY 578 GlyPheGlyGlyAspAlaPheIleAsnIleLysPheLeuTTPMetIleArgIleSerGluGlyIle 597
Db 77 GGTTCGGTGGTGTATGATATTAAGTTTATTAAGTTTATTAAGTTTATTAAGTTTATTAAGTTT 136
QY 598 TTPProGluGlu 601

```

Db 137 TGGCCAGAGAA 148

RESULT 9
CNS061LW 950 bp DNA linear GSS 17-JUN-2001
LOCUS T3 end of clone AT0AA009H03 of library AT0AA from strain CBS 4311
DEFINITION Saccharomyces servazzii, genomic survey sequence.
ACCESSION AL403626
VERSION AL403626.1 GI:12164096
KEYWORDS GSS.
SOURCE Saccharomyces servazzii
ORGANISM Saccharomyces servazzii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 950)
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
Maupertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P., and Weissenbach, J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584717
PUBMED 11152876
REFERENCE 2 (bases 1 to 950)
AUTHORS Casaregola, S., Lepingle, A., Bon, E., Neuveglise, C., Nguyen, H.,
Artiguenave, P., Wincker, P., and Gaillardin, C.
TITLE Genomic exploration of the hemiascomycetous yeasts: 7.
Saccharomyces servazzii
JOURNAL FEBS Lett. 487 (1), 47-51 (2000)
MEDLINE 20584717
PUBMED 11152882
REFERENCE 3 (bases 1 to 950)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
anastasi, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1..950
/organism="Saccharomyces servazzii"
/mol_type="genomic DNA"
/strain="CBS 4311"
/db_xref="taxon:27293"
/clone_lib="AT0AA009H03"
/clone_lib="AT0AA"
/note="end : T3"
misc_feature
95..9500
/note="similar to Saccharomyces cerevisiae ORF YGL022w [STM3 : oligosaccharyl transferase subunit]
1 putative frameshift(s)"
evidence=not_experimental
misc_feature
complement(837..9548)
/note="similar to Saccharomyces cerevisiae ORF YGL020c [hypothetical protein]"
evidence=not_experimental
ORIGIN
Alignment Scores:
Pred. No.: 6,88e-14 Length: 950
Score: 24.00 Matches: 24

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.34% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-6 (1-718) x CNS061LW (1-950)

QY 578 GlyPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIle 597
DB 77 GGTTCGGTGGTGATGATATTAATAAGTTTTTATGCGATGATCAGATTAGTGAAGGTATT 136

QY 598 TrpProGluGlu 601
DB 137 TGGCCAGAGAA 148

RESULT 10
CNS061LW 424 bp mRNA linear EST 29-JUN-2004
LOCUS EN15033.Sprime Exelixis FlyTag MN08 Bluescript Drosophila
DEFINITION melanogaster cDNA clone EN15033 5, mRNA sequence.
ACCESSION CO337666
VERSION CO337666.1 GI:49397941
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 424)
AUTHORS Nakanishi, M., Muzong, C., Peterson, E., Laufer, A., Leung, W., Piatt, D.
and Swimmer, C.
TITLE Exelixis FlyTag EST Project MN08 Library
JOURNAL Unpublished (2004)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd
Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: EN150 row: C column: 9
High quality sequence stop: 243.
FEATURES
source
1..424
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone_lib="EN15033"
/cell_line="mbn2"
/note="Vector: peluescript; Site 1: NotI; Site 2: XhoI;
oligodT primed from LSP induced mbn2 cell line."

ORIGIN
Alignment Scores:
Pred. No.: 5,07e-10 Length: 424
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 7 Gaps: 0

US-10-028-384-6 (1-718) x CO337666 (1-424)

QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
DB 56 TTTCATGCGCATCGTCCCTGGCTACATCATAGTAGTGGTGGCTGATCGTACGATAACGAG 115

RESULT 11
BM290810 576 bp mRNA linear EST 01-JUL-2002
LOCUS BM290810
DEFINITION EST577344 AvSG Amblyomma variegatum CDNA clone AVAAR06 5' end, mRNA
sequence.
ACCESSION BM290810

```

VERSION      BM290810.1  GI:21640777
KEYWORDS     EST.
SOURCE       Amblyomma variegatum
ORGANISM     Amblyomma variegatum
REFERENCE    1 (bases 1 to 576)
AUTHORS      Nene,V., Lee,D., Quackenbush,J., Skilton,R., Mwaura,S.,
              Gardner,M.J. and Bishop,R.
TITLE        AVGI, an index of genes transcribed in the salivary glands of the
              ixodid tick Amblyomma variegatum
JOURNAL      Int. J. Parasitol. 32 (12), 1447-1456 (2002)
MEDLINE      22281296
PUBMED       12392910
COMMENT      Contact: Vish Nene
              Parasite Genomics Group
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-510-5968
              Fax: 301-838-0208
              Email: nene@tigr.org
              Seq primer: M13 reverse.
              Location/Qualifiers
                1..576
                /organism="Amblyomma variegatum"
                /mol_type="mRNA"
                /db_xref="taxon:34610"
                /clone="AVAR06"
                /tissue_type="Salivary glands"
                /dev_stage="Adult"
                /lab_host="E.coli strain DH10B-Tona"
                /clone_lib="AVSG"
                /note="Vector: pCMV-SPORT6.1; Salivary glands were
              dissected on day five after initiation of feeding. Total
              RNA was prepared using acid guanidinium
              thiocyanate-phenol-chloroform extraction. The cDNA library
              was custom prepared by Invitrogen Corporation. Briefly,
              first strand cDNA was primed using oligo(dT) containing a
              NotI site. Size fractionated double stranded cDNA was
              ligated to EcoRV-NotI cleaved vector and electroporated
              into E.coli."
              ORIGIN
                Alignment Scores:
                Pred. No.: 6,88e-10 Length: 576
                Score: 20.00 Matches: 20
                Percent Similarity: 100.00% Conservative: 0
                Best Local Similarity: 100.00% Mismatches: 0
                Query Match: 2.79% Indels: 0
                DB: 4 Gaps: 0
                US-10-028-384-6 (1-718) x BM290810 (1-576)
                Qy 149 PheilealalevalProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
                Db 482 TTCATAGCCATTGTGCTGGTACATCAGCCGTTTCAGTGGCAGGAGTACGACAAATGAA 541
                RESULT 12
                A1133998 585 bp mRNA linear EST 19-APR-2001
                LOCUS GH11327.5prime GH Drosophila melanogaster head p0R2 Drosophila
                DEFINITION melanogaster cDNA clone GH11327 5prime, mRNA sequence.
                ACCESSION A1133998
                VERSION A1133998.1 GI:3626556
                KEYWORDS EST.
                SOURCE Drosophila melanogaster (fruit fly)
                ORGANISM Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                Phlebotominae; Drosophilidae; Drosophila.
                REFERENCE 1 (bases 1 to 585)
                AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
                Lewis,S. and Rubin,G.M.

```

```

TITLE      BGGP/HIMI Drosophila EST Project
JOURNAL     Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BGGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit P element sequence 1(3)j2D9
            Plate: 113 row: C column: 3
            High quality sequence stop: 470.
            Location/Qualifiers
              1..585
              /organism="Drosophila melanogaster"
              /mol_type="mRNA"
              /db_xref="taxon:7227"
              /clone="GH11327"
              /sex="male and female"
              /dev_stage="adult"
              /lab_host="PH5 - alpha"
              /clone_lib="GH Drosophila melanogaster head p0R2"
              /note="Organ: head; Vector: p0R2; Site 1: EcoRI; Site 2:
            XhoI; Sized fractionated cDNAs were directly ligated into
            p0R2. Plasmid cDNA library."
            ORIGIN
              Alignment Scores:
              Pred. No.: 6.99e-10 Length: 585
              Score: 20.00 Matches: 20
              Percent Similarity: 100.00% Conservative: 0
              Best Local Similarity: 100.00% Mismatches: 0
              Query Match: 2.79% Indels: 0
              DB: 1 Gaps: 0
              US-10-028-384-6 (1-718) x A1133998 (1-585)
              Qy 149 PheilealalevalProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
              Db 506 TTCATGCCATTCGCTGGTACATCAGTACATCAGTGGTGGTGGTACGATACGATACGAG 565
              RESULT 13
              BM638349 594 bp mRNA linear EST 26-FEB-2002
              LOCUS BM638349
              DEFINITION 17000687566421 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
              ACCESSION BM638349
              VERSION BM638349.1 GI:18937860
              KEYWORDS EST.
              SOURCE Anopheles gambiae (African malaria mosquito)
              ORGANISM Anopheles gambiae
              Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
              Anopheles.
              REFERENCE 1 (bases 1 to 594)
              AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
              Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
              TITLE CelerA Anopheles gambiae EST project
              JOURNAL Unpublished (2002)
              COMMENT Contact: Holt R.A.
              CelerA Genomics
              45 W. Gude Dr., Rockville, MD 20850, USA
              Tel: 2404533151
              Fax: 2404534580
              Email: HoltR@celerA.com
              Plate: NU01003CV8 row: O column: 23
              Seq primer: M13 Reverse.
              Location/Qualifiers
                1..594
                /organism="Anopheles gambiae"
                /mol_type="mRNA"
                /strain="RSP-ST (Reduced susc. to Permethrin - std.
                chromosome)"
                /db_xref="taxon:7165"

```



```

/clone="19600449630257"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="A.Gam.ad.cdna1"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3' Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Alignment Scores:
Pred. No.: 7,1e-10 Length: 594
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x BM638349 (1-594)
Qy 149 PheilleAlaIleValProGlyTyrlleSerArgSerValAlaGlySerTyAspAsnGlu 168
Db 320 TTCATCGGATCGTGGCCGGCTACACGCCGCTCGTGGCGGCTCGTACGATACCGAG 379

RESULT 14
Bi170396
LOCUS
DEFINITION
REI1825.Sprime RE Drosophila melanogaster normalized Embryo pF1c-1
Drosophila melanogaster cDNA clone REI1825 5 similar to OstStt3:
FBan007748 'enzyme' located on: 3R 96B16-96B17; 04/12/2001, mRNA
sequence.
Bi170396
ACCESSION
Bi170396.1 GI:14636203
VERSION
EST.
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 598)
Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
Rubin,G.M.
BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 04/12/2001 hit P element 1(3)j2D9:
1(3)j2D9 A026308 inserted at base 292 5' end of P element Inverse
PCR: 03/18/2001
Plate: RE:118 row: C column: 1
High quality sequence stop: 547.
Location/Qualifiers
1..598
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pF1c-1"

FEATURES
source

```

```

/notes="Organ: embryo; Vector: pF1c1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.: 7,1e-10 Length: 598
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.79% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-6 (1-718) x Bi170396 (1-598)
Qy 149 PheilleAlaIleValProGlyTyrlleSerArgSerValAlaGlySerTyAspAsnGlu 168
Db 483 TTCATCGGATCGTGGCCGGCTACATCATAGTGGTGGCTGGATCGTACGATACCGAG 542

RESULT 15
Bi163592
LOCUS
DEFINITION
RE03026.Sprime RE Drosophila melanogaster normalized Embryo pF1c-1
Drosophila melanogaster cDNA clone RE03026 5 similar to OstStt3:
FBan007748 'enzyme' located on: 3R 96B16-96B17; 04/11/2001, mRNA
sequence.
Bi163592
ACCESSION
Bi163592.1 GI:14629398
VERSION
EST.
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 600)
Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
Rubin,G.M.
BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 04/11/2001 hit P element 1(3)j2D9:
1(3)j2D9 A026308 inserted at base 292 5' end of P element Inverse
PCR: 03/17/2001
Plate: RE:30 row: C column: 2
High quality sequence stop: 455.
Location/Qualifiers
1..600
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pF1c-1"
/notes="Organ: embryo; Vector: pF1c1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN

```

Alignment Scores:

| Alignment Scores: | | |
|------------------------|----------|-----------------|
| Pred. No.: | 7,176-10 | Length: 600 |
| Score: | 20.00 | Matches: 20 |
| Percent Similarity: | 100.00% | Conservative: 0 |
| Best Local Similarity: | 100.00% | Mismatches: 0 |
| Query Match: | 2.79% | Indels: 0 |
| DB: | 4 | Gaps: 0 |

US-10-028-384-6 (1-718) x BI163592 (1-600)

Qy 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
| | | | |
Dd 499 TTCATGCCCATCGTGCCCTGGCTACATCATAGTCGGTGGCTGGATCGTACGATAACGAG 558
| | | | |

Search completed: December 15, 2004, 06:14:10
Job time : 5570.03 secs

1297

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 7954.29 Seconds
(without alignments)
4268.640 Million cell updates/sec

Title: US-10-028-384-6
Perfect score: 3761
Sequence: 1 MGSDFSCVLSVFOTILKLV.....TRSTKTRBSIKRPELGLRV 718

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DSV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app_query.fasta_1.3740
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -NATX=Biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORV=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn 1 1 20262 @runat_14122004_131519_8017 -NCFU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -YGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 3761 | 100.0 | 2733 | 6 | AX799086 Sequence |
| 2 | 3761 | 100.0 | 2733 | 8 | D28952 Saccharomyc |
| 3 | 3751 | 99.7 | 3633 | 8 | SCYGL022W |
| 4 | 3488 | 92.7 | 2000 | 6 | AX595714 Sequence |

| 5 | 3488 | 92.7 | 2000 | 6 | AX819898 |
|----|--------|------|--------|----|-------------|
| 6 | 3488 | 92.7 | 2000 | 6 | AX830928 |
| 7 | 2969 | 78.9 | 110000 | 8 | CR380947_0 |
| 8 | 2905 | 77.2 | 110000 | 8 | CR382123_09 |
| 9 | 2905 | 77.2 | 110000 | 8 | CR382123_10 |
| 10 | 2901 | 77.1 | 300029 | 8 | AE016903 |
| 11 | 2708 | 72.8 | 2256 | 6 | AX489023 |
| 12 | 2708 | 72.0 | 110000 | 8 | CR382134_03 |
| 13 | 2495.5 | 66.4 | 110000 | 8 | CR382128_21 |
| 14 | 2237 | 55.5 | 40907 | 8 | SPBIC1271 |
| 15 | 2224.5 | 59.1 | 5048 | 8 | AX801532 |
| 16 | 2180.5 | 58.0 | 94169 | 8 | AX842635 |
| 17 | 2027.5 | 53.9 | 2417 | 3 | AF132552 |
| 18 | 2027.5 | 53.9 | 2417 | 3 | AX799088 |
| 19 | 2027.5 | 53.9 | 2699 | 6 | CQ589353 |
| 20 | 2016.5 | 53.6 | 4236 | 10 | BC052433 |
| 21 | 2013.5 | 53.5 | 2710 | 6 | AX799084 |
| 22 | 2012.5 | 53.5 | 2481 | 6 | AX799082 |
| 23 | 2012.5 | 53.5 | 2481 | 9 | AY074880 |
| 24 | 1966 | 52.0 | 2567 | 5 | BC063234 |
| 25 | 1956 | 52.0 | 2608 | 5 | BC046072 |
| 26 | 1955.5 | 52.0 | 4922 | 6 | CQ589352 |
| 27 | 1955.5 | 52.0 | 162921 | 3 | AC007853 |
| 28 | 1955.5 | 52.0 | 181132 | 3 | AC008206 |
| 29 | 1955.5 | 52.0 | 227219 | 3 | AE003750 |
| 30 | 1949 | 51.8 | 2118 | 12 | BT008132 |
| 31 | 1946 | 51.7 | 2118 | 9 | BT007100 |
| 32 | 1946 | 51.7 | 2450 | 9 | BC048348 |
| 33 | 1946 | 51.7 | 2458 | 6 | CQ834714 |
| 34 | 1946 | 51.7 | 2472 | 6 | CQ727867 |
| 35 | 1946 | 51.7 | 2516 | 9 | BC020965 |
| 36 | 1945 | 51.7 | 2293 | 9 | HUMTWC |
| 37 | 1945 | 51.7 | 2736 | 10 | BC037612 |
| 38 | 1945 | 51.7 | 3094 | 6 | AX799090 |
| 39 | 1945 | 51.7 | 3094 | 10 | MUSRPR |
| 40 | 1940.5 | 51.6 | 2481 | 5 | BC067313 |
| 41 | 1937 | 51.5 | 3666 | 10 | AK129027 |
| 42 | 1934 | 51.4 | 2898 | 9 | HSMB09254 |
| 43 | 1928.5 | 51.3 | 2435 | 6 | CQ834710 |
| 44 | 1928.5 | 51.3 | 75650 | 2 | AC018145 |
| 45 | 1920 | 51.1 | 2472 | 6 | AX799092 |

ALIGNMENTS

RESULT 1
AX799086
LOCUS AX799086 2733 bp DNA linear PAT 08-OCT-2003
DEFINITION Sequence 5 from Patent WO03054008.
ACCESSION AX799086
VERSION AX799086.1 GI:37605060
KEYWORDS
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

Location/Qualifiers

1. 2733

/organism="Saccharomyces cerevisiae"

/mol_type="genomic DNA"

/db_xref="taxon:4932"

/note="STT3, Acc# D28952"

ORIGIN

Alignment Scores:

Pred. No.: 7.58e-305

Score: 3761.00

Length: 2733

Matches: 718

Qy 701 SerSerThrLysThrArgAspSerIleLysArgProGluLeuGlyLeuArgVal 718
 Db 2319 TCCTTACGAAACCAAGAGGTCCATTAAGAGACCTGAATTAGGCTTGAGAGTC 2372

RESULT 2
 YSCSTT3 2733 bp DNA linear PLN 18-JAN-2003
 LOCUS Saccharomyces cerevisiae gene for STT3 protein, complete cds.
 DEFINITION D28952
 ACCESSION D28952
 VERSION D28952.1 GI:468493
 KEYWORDS
 SOURCE Saccharomyces cerevisiae (baker's yeast)
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1
 AUTHORS Yoshida S., Ohya Y., Nakano A. and Anraku Y.
 TITLE STT3, a novel essential gene related to the PKC1/STT1 protein kinase pathway, is involved in protein glycosylation in yeast
 JOURNAL Gene 164 (1), 167-172 (1995)
 MEDLINE 96060859
 PUBMED 7590309
 REFERENCE 2 (bases 1 to 2733)
 AUTHORS Yoshida S.
 TITLE Direct Submission
 JOURNAL Submitted (14-MAR-1994) Satoshi Yoshida, Faculty of Science, University of Tokyo, Department of Biology; 7-3-1, Hongo, Bunkyo-Ku, Tokyo 113, Japan (Tel:81-3-3812-2111 (ex.4461), Fax:81-3-3812-4929)

FEATURES
 Location/Qualifiers
 1..2733
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /strain="S288C"
 /db_xref="taxon:4932"
 /clone="pST3-105"
 /clone_lib="library of T. Yoshihisa and Y. Anraku"
 gene 219..2375
 /gene="STT3"
 CDS 219..2375
 /gene="STT3"
 /codon_start=1
 /product="STT3 protein"
 /protein_id="BAA06079.1"
 /db_xref="GI:603808"

/translation="MGSDRSCLVSVFQTLKLVIVFPAISRLFAVIFKESIIH
 EFDPNRYATKLVNNSFYKFLNPFDDRTWYPLRGVGGTLYPLGLTTSFIFWALR
 NMLGLPIDIRNVCVLPAPLFGVTAWATYFTKEIKDASAGLLAAGFIAIVPGYISRS
 VAGSYDNEAIAITLLMTVMFKQKGTGIMHATCAALFYFMYSAWGVYFVFNLI
 PLHFVLLILMGYSKLSAYTTWTAIGTVASMOIPFVGFLPIRNDHMAALGVFGLI
 QIVAFGDFVKGQISTAKPKVIMMVSILFLILVGVGLSALTVMGLIAPWGPVFLWD
 TVYAKTHIPILASVSEHQVSPHAPFFDTHFLILWFLPAGVFLFLDLKDERVFIAYS
 VLCSYFAGVWVRMLTLFVPCVSAVALSKIFDYLDFKTSDDRIYAKPAALLAKLI
 VSGSFYILYLFVHSTWTRTAYSPSVLPSPVTPDGLALIDDFREAYVWLRNSD
 EDSKAAWMDYQIQGMADRTLVNDNTNNTHIAIVGKAMASPEEKSEYILKEHDV
 DYLVIFGLLIGFGDDINKFLMIRISEGIMPEIKERYFYTAEGEYRVVARASETM
 RNSLLYKMSYKQFPQLFNGQQTDRVQOMITPLDVPDLDPYDFVFTSENWVRVYQL
 KXDDAQRGLRDVGELTRSSTKTRRSIKPELGLRV"

ORIGIN

Alignment Scores:
 Pred. No.: 7,58e-305 Length: 2733
 Score: 3761.00 Matches: 718
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-028-384-6 (1-718) x YSCSTT3 (1-2733)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 |||||

Db 219 ATGGATCCGACCGGTCGTGTGTTTGTCTGTGTTTCAGACCATCTCAAGCTCGTCATC 278
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 |||||
 Db 279 TTCGCGGGGATTTTTTGGGGCTGCCATATCATCACGTTTGTGTCAGTCATCAAAATTGAG 338
 Qy 41 SerIleIleHisGluPheAspProTrrPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 |||||
 Db 339 TCTATTATCCATGAATTCGACCCCTGGTTCATATATAGGGCTACCAATATCTCGTCAC 398
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrrPheAspArgThrTrrPrrProLeuGlyArg 80
 |||||
 Db 399 AATTCGTTTACAAGTTTTTGAACCTGGTTTGACGACCGTACCTGGTACCCCTCGGAGG 458
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrrPheAla 100
 |||||
 Db 459 GTTACTGGAGGAGCTTTATATCTGGTTTGAAGCAGTAGTGGTTCATCTGGCAGCC 518
 Qy 101 LeuArgAsnTrrPLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 |||||
 Db 519 CTGCGCAACTGGTTGGGCTTGCCTTGCATCAGACATCAGAAACGTTTGTGTCTATTTCGCCA 578
 Qy 121 LeuPheSerGlyValThrAlaTrrPAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 |||||
 Db 579 CTATTTCTGGGGTCACCGCCTGGCGGACTTACGAATTTACGAAGAGATTAAAGATGCC 638
 Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 |||||
 Db 639 AGCGTGGGCTTTGGCTGCTGGTGTATATAGCCATTGTCGCCGTATATATCTAGATCA 698
 Qy 161 VallAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 |||||
 Db 699 GTGGCGGGCTCTACGATATATGAGCCATTGCCATTACACTATTAAATGGTCACATTTTCATG 758
 Qy 181 PheTrrPleLysAlaGlnLysThrGlySerIleMethHisAlaThrCysAlaAlaLeuPhe 200
 |||||
 Db 759 TTTTGGATTAGGCCCAAGAGACTGGCTCTATCATGACGCAAGCTGTGCAGCTTTTATTC 818
 Qy 201 TyrPheTyrMetValSerAlaTrrPglyTyrValPheIleThrAsnLeuIleProLeu 220
 |||||
 Db 819 TACTTCTACATGGTGTGGCTTGGGGTGGATACGTGTTCATCACCACCTTGATCCCACTC 878
 Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTrrThr 240
 |||||
 Db 879 CATGCTTTTGTGATTTTGTATGGCGCAGATATTCGCCAAACGTGATTTCTGCTACACC 938
 Qy 241 ThrTrrPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 |||||
 Db 939 ACTTGGTACGCTATTGGAACCTGTGCATCCATGCAGATCCCAATTTGTGCGGTTTCTACCT 998
 Qy 261 IleArgSerAsnAspHisMetAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 |||||
 Db 999 ATCAGGCTAACGACCACATGGCGCGCATGGGTGTTTTTCGGTTTGATTCAGATTGTGCGC 1058
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleIleMetMetVal 300
 |||||
 Db 1059 TTGCGTGACTTCGTGAAGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGTT 1118
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 |||||
 Db 1119 TCTCTGTTTTGTATCTTGGTCTTGGTGTGGTGTGGATCTTCTGCTTGAACATATATGGGG 1178
 Qy 321 LeuIleAlaProTrrPThrGlyArgPheTyrSerLeuTrrPAspThrAsnTyrAlaValIle 340
 |||||
 Db 1179 TTGATTGCGGCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1238
 Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrrPrrProAlaPhePhe 360
 |||||
 Db 1239 CACATTCCTATCATTCGCTCCGTTTCCGAACATCAACCCCGTTTCGTGGCGCGCTTCTTCTC 1298
 Qy 361 PheAspThrHisPheLeuIleTrrPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 |||||
 Db 1299 TTTGATACCCACATTTTTTGATCTGCTATCCCCCGCGGTGATTCTCTACTATTCTCTCGAC 1358

Db 1253 ATGGGATCCGACCGGCTGGTGTGTTTGTCTGTGTTTCAGACCACTCCTCAAGCTCGTCATC 1312
Qy 21 PheValAlaIlePheGlyValAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
Db 1313 TTCGGGGGCGATTTTGGGGCTGGCCATATCATCATCGTTTGTTCGAGTCATCAAAATTTGAG 1372
Qy 41 SerIleIleHisGluPheAspProIlePheAsnTyrArgAlaThrLysTyrLeuValAsn 60
Db 1373 TCTATTATCCATGAATTCGACCCCTGGTTCATATATAGGGCTACCAAAATATCTCGTCAAC 1432
Qy 61 AsnSerPheTyrLysPheLeuAsnTyrPheAspArgThrTyrProLeuGlyArg 80
Db 1433 AATTCTGTTTCAAGTTTTCAGCTGGTTTGACACCGTACCTGGTACCCCTCGGAAGG 1492
Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTyrHisAla 100
Db 1493 GTTACTGGAGGACATTTATATCTCTGGTTGTATGACGACGTAGTGGTTCATCTGGCACGCC 1552
Qy 101 LeuArgAsnTyrLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
Db 1553 CTGCGCACTGGTTGGGCTTGCCCATTTGACATCAGAAACGGTTTGTGTCTATTTCGGCCA 1612
Qy 121 LeuPheSerGlyValThrAlaTyrAlaThrTyrGluPheThrLysGluIleLysAspAla 140
Db 1613 CTATTTCTGGGGTCCACCGCTGGCGGACTTACGAAATTTACGAAAGAGATTAAAGATGCC 1672
Qy 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
Db 1673 AGCGTGGGCTTTGGCTGCTGGTTTATAGCCATGTCCCGGTTATATATCTAGATCA 1732
Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
Db 1733 GTGGCGGGGCTCTACGATAATCAGCGCAATGCCATTACACTATTAAATGGTCACTTCATG 1792
Qy 181 PheTyrIleLysAlaGlyLysThrGlySerIleMethHisAlaThrCysAlaAlaLeuPhe 200
Db 1793 TTTTGGATTAGGCCCAAAAGACTGGCTCTATCATATGACGACGCTGGCAGCTTTATTC 1852
Qy 201 TyrPheTyrMetValSerAlaTyrGlyTyrValPheIleThrAsnLeuIleProLeu 220
Db 1853 TACTTCTACATGGTGTGGCTGGGTGGATACGCTGTTCATCAACCACTTGTATCCCACTC 1912
Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
Db 1913 CATGCTTTTCTGCTGATTTTGATGGCGCAGATATTCGTCGCAACCTGATTCGCTACACC 1972
Qy 241 ThrTyrPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
Db 1973 ACTTGGTACGCTATTGGGAACGTGTGCATCCATGCAGATCCCATTTGTGGTTCCTACCT 2032
Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
Db 2033 ATCAGGCTTACGACACCATGGCCGCAATGGGTGTTTCGGTTTGATTCAGATTGTCGCC 2092
Qy 281 PheGlyAspPheValIleGlyLysThrAlaLysPheLysValIleMetMetVal 300
Db 2093 TTCCGTGATCTTCGTGAAGGGCCAAATCAGCACAGCTAAGTTTAAAGTCATCATGATGTT 2152
Qy 301 SerLeuPheLeuIleLeuValIleGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
Db 2153 TCTCTGTTTGTATCTTGGTCTTGGTGTGTCGACCTTTCGCTTTCGCTTTCAGTTCATGCGG 2212
Qy 321 LeuIleAlaProTyrThrGlyArgPheTyrSerLeuTyrAspThrAsnTyrAlaLysIle 340
Db 2213 TTGATTGCCCCCTGGCATGTTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 2272
Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTyrProAlaPhePhe 360
Db 2273 CACATTCCTATCTTGGCTTCGTTCCGAACATCACCCGTTTCTGGTGGCCCGCTTCTTC 2332
Qy 361 PheAspThrHisPheLeuIleTyrPhePheProIleGlyValPheLeuLeuPheLeuAsp 380

RESULT 4
AX595714
LOCUS

AX595714

2000 bp

DNA

linear

PAT 14-FEB-2003

Db 2333 TTTGATACCCACTTTTGTGATCTGGCTATTCCCGCGGGTGTATTCCTACTATTCTCTCGAC 2392
Qy 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
Db 2393 TTGAAAACGACGACACGTTTGTTCATCGCTTACTCGGTTCTGTGTTCTACTTTGCGGT 2452
Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
Db 2453 GTTATGGTTAGATTGATTGTTGACTTTGACACCACTCATCTGTGTGTCCCGCGCGTCGA 2512
Qy 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
Db 2513 TTGTCCAAGATATTGACATCTACTTGGATTTCACAGACAAGTCACGCAATACGCCATC 2572
Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
Db 2573 AAACCTCGGCACCTACTGGCCAAATGATTGTTTCCGGATCATCTCATCTTTTATTGTAT 2632
Qy 461 LeuPheValPheHisSerThrTyrValThrArgThrAlaTyrSerSerProSerValVal 480
Db 2633 CTTTTCTGCTTCCATTTCTACTTGGTTAAACAAGAACTGCATCTCTCTCTCTCTCTCT 2692
Qy 481 LeuProSerGlnThrProAspGlyLysLeuAlaIleLeuLeuAspPheArgGluAlaTyr 500
Db 2693 TTGCCATCACAAACCCAGATGGTAAATTTGGCGTTGATCGACACTTCAGGGAAGCGTAC 2752
Qy 501 TyrTyrLeuArgMetAsnSerAspGluAspSerLysValAlaAlaIleTyrPheTyrGly 520
Db 2753 TATTGGTTTAAAGTGAATCTCTGATGAGGACAGTAAAGTTGCACGGTGGGATTACGGT 2812
Qy 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTyrAsnAsn 540
Db 2813 TACCAAAATTTGGTGGCATGGCAGACAGAACCACTTTTGTAGTAAACAACACGTTGGAACAAT 2872
Qy 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
Db 2873 ACTCACATCGCATCGTTGGTAAAGCATGCTTCCCTGAAGAGAAATCTTACGAAAT 2932
Qy 561 LeuLysGlnHisAspValAspTyrValLeuValIlePheGlyLysLeuIleGlyPheGly 580
Db 2933 CTAAGAAGCATGATGATGTCGATTTATGCTTGTGTCATCTTTGGTGTCTAAATGGGTTTGT 2992
Qy 581 GlyAspAspIleAsnLysPheLeuTyrMetIleArgIleSerGluGlyIleTyrProGlu 600
Db 2993 GGTGATGATCATCAACAAATTTCTTGTGGATGATCAGAAATTTAGCGAGGATCTGGCCAGAA 3052
Qy 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla 620
Db 3053 GAGATAAAGAGCGTGTATTTCTATACCGCAGAGGAGAAATACAGAGTAGATGCAAGGCT 3112
Qy 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
Db 3113 TCTGAGACCATGAGAACTCGCTACTTTACAAGATGCTCTACAAGATTTCCACAAATTA 3172
Qy 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
Db 3173 TTCAATGGTGCCCAAGCCACTGACAGAGTGGTCAACAAATGATCACCAATTAGACGTC 3232
Qy 661 ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTyrMetValArgIleTyr 680
Db 3233 CCACCATTTAGTACTTCTGACGAGATTTTACTTTCCGAAAACTGGATGGTTAGAAATAT 3292
Qy 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
Db 3293 CAATTGAAGAAGGATGATGCCAAGGTAGAACTTTGAGGACGTTGGTGGATTAAACAGG 3352
Qy 701 SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
Db 3353 TCTTCTAGAAACCAAGAGGTCCATTAAGAGAGACCTGAATTAGCTTTGAGATC 3406

DEFINITION Sequence 1368 from Patent EPI258494.
 ACCESSION AX595714
 VERSION AX595714.1 GI:28396965
 KEYWORDS Saccharomyces cerevisiae (baker's yeast)
 SOURCE Saccharomyces cerevisiae
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1
 AUTHORS Bauer A., Gavin, A.C., Grandi, P., Krause, R., Kruse, U., Kuster, B.,
 Martioch, M., Schultz, J. and Superti-Furga, G.
 TITLE Multiprotein complexes from eukaryotes
 JOURNAL Patent: EP 1258494-A 1368 20-NOV-2002;
 CELLZONE AG (DE)

FEATURES
 source Location/Qualifiers
 1..2000
 /organism="Saccharomyces cerevisiae"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4932"

ORIGIN

Alignment Scores:
 Pred. No.: 3,578-282 Length: 2000
 Score: 3488.00 Matches: 665
 Percent Similarity: 99.85% Conservative: 0
 Best Local Similarity: 99.85% Mismatches: 1
 Query Match: 92.74% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-6 (1-718) x AX595714 (1-2000)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 DB 1 ATGGGATCCGACCGGTCGTGTGTTTGTCTGTGTTTTCAGACCATCTCTCAAGTCGTCTATC 60
 QY 21 PheValAlaIlePheGlyAlaAlaIleSerArgLeuPheAlaValIleIlePheGlu 40
 DB 61 TTCGTGGCGGATTTTGGGGTGGCCATATCATACGCTTTGTTGGAGTCATCAAAATTTGAG 120
 QY 41 SerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 DB 121 TCTATTATCCATGAATTCGACCCCTGGTTCAATTATAGGCTTACCAAAATATCTGTCAC 180
 QY 61 AsnSerPheTyrLysPheLeuAsnTTPheAspAspAtcThrTyrPheProLeuGlyArg 80
 DB 181 AATTTCGTTTACAAATTTTGAATCTGTTGAGCGGACCGTACCTGGTACCCCTCGGAAGG 240
 QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
 DB 241 GTTACTGGAGGGACTTTATATCTGCTGTTTGGTACGACTAGTGGCTTCATCTGGCAGCC 300
 QY 101 LeuArgAsnTTPheGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 DB 301 CTGGCAACCTGGTGGCTTGGCCATTCATCATCAGAAAGTTTGTGCTATTTCGGCCA 360
 QY 121 LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 DB 361 CTATTTTCTGGGGTACCGGCTGGGAGCTTACGAATTTACGAAAGAGATTAAAGATGCC 420
 QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 DB 421 AGCGCTGGGCTTTGGCTGCTGTTTATAGCCATTGTCCTCCGGTTATATATCTAGATCA 480
 QY 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 DB 481 GTGGCGGGGCTTACCATATAGAGCCATTGGCCATTACACTATTAAATGGTCACTTTTCATG 540
 QY 181 PheTTPheLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 DB 541 TTTTGGATTAAAGCCCAAGAGCTGCTCTATCATGACGCGACCGTGTGACGCTTTATTC 600
 QY 201 TyrPheTyrMetValSerAlaTTPGlyTyrValPheIleThrAsnLeuIleProLeu 220

DB 601 TACTTCTCATCATGGTGTGCGGCTTGGGGTGGATACGCTGTTTCATCACCACACTTGTATCCCACTC 660
 QY 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 DB 661 CANGTCCTTTTGTCTGATTTTGTATGGGCGAGATATTGCTCCAAACTGTATTTCTGCTACACC 720
 QY 241 ThrTTPyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 DB 721 ACTTGGTACGCTATTGGAACTGTTGCATCCATGAGATCCCATTTGCGGTTTCTTACCT 780
 QY 261 IleAspSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 DB 781 ATCAGGCTCTACGACACACATGCGCGCATTTGGGTGTTTCGTTTTCATTCAGATTGCGC 840
 QY 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 DB 841 TTCGGTGCATTCGTGAGGGGCCAAATCAGACAGCTAAGTTTAAAGTCATCATGATGGTT 900
 QY 301 SerLeuPheLeuLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 DB 901 TCTCTGTTTTCATCTTGGTCTTGGTGTGGTGTGGACTTTCGCTTTCAGCTATATGGGG 960
 QY 321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle 340
 DB 961 TTGATTGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1020
 QY 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe 360
 DB 1021 CACATTCCTATCATTCGCTCGGTTCCGAACATCAACCGCTTTCGTCGCGCTTCTTC 1080
 QY 361 PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 DB 1081 TTTGATACCCACTTTTTCATCTGCTATTCGCCCGGTGTATTTCTTACTATTCTTCGAC 1140
 QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
 DB 1141 TTGAAAGCGAGCAGCGTTTGTTCATCGCTTACTCCGTTCTGTGTTGCTACTTTCGGGT 1200
 QY 401 ValMetValArgLeuMetLeuLeuThrProValIleCysValSerAlaAlaValAla 420
 DB 1201 GTTATGGTTAGATTGATGTTGACTTTGCACACAGTCTCTGTGTGTCGCGCGCGTGC 1260
 QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 DB 1261 TTGTCAGAGATATTTCATCTACTCTGGATTTTCAGACAGTGCACCGCAATACGCCATC 1320
 QY 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 DB 1321 AAACCTGGCGGCACTACTGCGCAAAATTTGATTGTTTCCGGATCATTCATCTTTTATTTGT 1380
 QY 461 LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal 480
 DB 1381 CTTTTCGTCCTTCATCTCTCTGGTACAGAACTGCATCTCTCTCTCTCTCTCTCTCTCT 1440
 QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspPheArgGluAlaTyr 500
 DB 1441 TTGCCATCACAAACCCAGATGGTAAATTTGGCGTTTGTATCGACGACTTCAGCGAAGCGTAC 1500
 QY 501 TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPAspTyrGly 520
 DB 1501 TATTGGTTAAGATGAATCTCTGTAGAGACAGTAAAGTTGCAGCGTGTGGGATTCAGGT 1560
 QY 521 TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTTPAsnAsn 540
 DB 1561 TACCAAAATTTGGTGGATGGCAGACAGACCACTTTTGTCTGATATACACACAGCTGGAACA 1620
 QY 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
 DB 1621 ACTCATATCGCCATCGTTGGTAAAGCCATGGCTTCCCTGAGAGAGAAATCTTACGAAAT 1680
 QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIleGlyPheGly 580
 DB 1681 CTAAAAGAGCATGATGATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740

Db 1321 AACCTGGGCACTACTGCGCAAAATTGATTGTTCCGGATCATCTTATTTTATTTGAT 1380
 Qy 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
 Db 1381 CTTTTCGTCTTCATCTTCTCTGGGTAAACAGAACTGCATCTCTCTCTCTCTCTCTCT 1440
 Qy 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeu 500
 Db 1441 TTGCATCAACAACCCAGATGTAATTTGGCTTGCATCGACGATTCAGGAGCGTAC 1500
 Qy 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGly 520
 Db 1501 TATTGGTTAGAACTCTGATGAGGACAGTAAGTTGCGGTGGGTGGGATTTACGGT 1560
 Qy 521 TyrGlnIleGlyMetAlaAspArgThrThrLeuValAlaAspAsnAsnThrTrpAsnAsn 540
 Db 1561 TACCAATTTGGTCATGGCAGACAGAACCCATTTAGTCGATAACAACACGTTGGAACAAT 1620
 Qy 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
 Db 1621 ACTCACATCGCCATCGTTGTAAAGCCATGCTTCCCTGAGAGAAATCTTACGAAAT 1680
 Qy 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1681 CTAAGAGCATGATGTCGATTTATGCTTGTGTCATCTTTGGTGGTCTAATTTGGTGGT 1740
 Qy 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1741 GGTGATGACATCAACAATTTCTGTGATGATCAGAAATTTAGCGAGGAAATCTGGCCAGA 1800
 Qy 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla 620
 Db 1801 GAGATAAAGAGCGTGATTTCTATACCGCAGAGGAGAAATACAGACTAGATCAAGGCT 1860
 Qy 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
 Db 1861 TCTGAGACCATGAGAACTCGCTACTTTACAGATGCTCTCAAGATTTCCCAAAATTA 1920
 Qy 641 PheAsnGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
 Db 1921 TTCAATGTGGCCAAAGCACTGACAGAGTGGTCAACAAATGATCACACCAATAGACGTC 1980
 Qy 661 ProProLeuAspTyrPhe 666
 Db 1981 CCACATTAGACTACTTC 1998
 RESULT 6
 AX830928 2000 bp DNA linear PAT 12-DEC-2003
 LOCUS Sequence 1648 from Patent WO03072602.
 DEFINITION AX830928
 ACCESSION AX830928
 VERSION AX830928.1 GI:39839809
 KEYWORDS Saccharomyces cerevisiae (baker's yeast)
 SOURCE Saccharomyces cerevisiae
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE
 AUTHORS Bauer, A., Gavin, A. C., Superti-Furga, G., Kuster, B., Schultz, J.,
 Marzioch, M., Grandi, P., Krause, R., Kruse, U., Merino, A., Bauch, A.,
 Michon, A. M., Leutwein, C. and Rick, J.
 TITLE Protein complexes and methods for their use
 JOURNAL Patent: WO 03072602-A 1648 04-SEP-2003;
 CELLZONE AG (DE)
 FEATURES
 source
 1. 2000
 Location/Qualifiers
 /organism="Saccharomyces cerevisiae"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4932"
 ORIGIN
 Alignment Scores: 3.57e-282 Length: 2000
 Pred. No.:

Score: 3488.00 Matches: 665
 Percent Similarity: 99.85% Conservative: 0
 Best Local Similarity: 99.85% Mismatches: 1
 Query Match: 92.74% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-6 (1-718) x AX830928 (1-2000)
 Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 1 ATGGGATCCGACCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 60
 Qy 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 61 TTCTGGTGGCGAATTTTGGGGTGGCCATATCATCAGCTTTGTTTGCAGTCATCAAAATTGAG 120
 Qy 41 SerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 Db 121 TCTATATCCATGATTCGACCCCTGGTTCATTTATAGGCTACCAATATCTCGTCAAC 180
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
 Db 181 AATTCGTGTTTACCAAGTTTTCGAACCTGTTTGCACACCTACCTGGTACCCCTCGGAAGG 240
 Qy 81 ValThrGlyLysThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTyrHisAla 100
 Db 241 GTTACTGGAGGACATTTTATATCTGGTTCATGATGACGACTAGTGGTTCATCTGGCAGCC 300
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 301 CTGCGCAACTGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 Db 361 CTATTTCTGGGGTCAACCGCTGGGGGACTTACGAATTTAGGAAGAGATTAAGATGCC 420
 Qy 141 SerAlaGlyLeuLeuAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 421 AGCGCTGGGCTTTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 480
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 481 GTGGGGGGTCTCTACGATAATGAGGCCATTCGCCATTACACTATTATTAAGTCACTTCATG 540
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
 Db 541 TTTTGGATTAAAGGCCCAAAAGACTGGCTCTATCATGACGCAACGCTGTGAGCTTTATTC 600
 Qy 201 TyrPheTyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 601 TACTTCTACATGGTCTGGGCTGGGTGGATACGCTGTTTCATCCCAACTTGATCCCACTC 660
 Qy 221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 661 CATGTCTTTTCTCTGATTTTGTATGGGCGAGATATTTCGTCAAACTGTTATTCGCTACACC 720
 Qy 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 721 ACTTGGTACGCTATTGGAACTGTTGTCATCCATGACAGATCCCATTTGCGGTTTCTACCT 780
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 781 ATCAGGTCTTAACGACCACTACGCGCATTTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 840
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 841 TTCTGGTGGCTCTCGTGAAGGCCCAATCAGCAGACTAAGTTTAAAGTCATCATGATGGTT 900
 Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 901 TCTCTGTTTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 Qy 321 LeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340


```

Db 961 TTGATGGCCCTTGACCTGGTAGATTTTATTGGTATGGGATACCAACTACGCAAGATC 1020
QY 341 HisIleProIleAlaSerValSerGluHisGlnProValSerTyrProAlaPhePhe 360
Db 1021 CACATTCCTATCATTCGCTCGCTTTCCGACATCAACCCGTTTCGTGGCCGCTTCTTC 1080
QY 361 PheAspThrHisPheLeuIleThrLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
Db 1081 TTGTATACCATCTTTTGATCGSGTATTCGCCGCGGTGATTTCCTACATATTCCTGCAG 1140
QY 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
Db 1141 TTGAAGACGACGACGCTTTTGTGCTTACCTCGTTCGTGCTGCTACTTTCGCGGT 1200
QY 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValala 420
Db 1201 GTTATGGTTAGATTGATGTGATCTTGACACGAGTCATCTGTGTGTCGCGCGCGTCGCA 1260
QY 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
Db 1261 TTGTCCACGATATTTGACATCTACCTGGATTTCAAGACAGTACCGCAAAATACGCCATC 1320
QY 441 LysProAlaAlaLeuAlaLysIleValSerGlySerPheIlePheTyrLeuTyr 460
Db 1321 AAACCTGCGGCACTACTGCCAAATGTGATTTGCCGATCATTCATCTTTTATTGTAT 1380
QY 461 LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal 480
Db 1381 CTTTTCGTCTTCCATCTACTTGGTAAACAGACGTCATCTCTCTCTCTCTCTCTCTCT 1440
QY 481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
Db 1441 TTGCCATCAACAAACCCAGATGGTAAATTTGGCTTTGATCGACGACTTCAGGAGCGCTAC 1500
QY 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTyrPheTyrGly 520
Db 1501 TATTGGTTAAGAAATGAACCTCTGATGAGCAGATGAAGTTGACGCGTGTGGGATTACGGT 1560
QY 521 TyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
Db 1561 TACCAAAATGGTGGATGTCAGACAGAACCACTTTAGTCGATACACACAGCTGGAAACAT 1620
QY 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle 560
Db 1621 ACTCACATCGCATCCTTGGTAAAGCCATGCTTCCCTGAAGAGAAATCTTACGAAAT 1680
QY 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyLeuIleGlyPheGly 580
Db 1681 CTAAAGACGATGATGCGATATGCTTGTGTCATCTTTGTGTCTCTAAATGGGTTTGGT 1740
QY 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTyrProGlu 600
Db 1741 GGTGATGATCAACAAATCTTTGTGGATGATCAGAAATAGCAGGGAATCTGCCAGAA 1800
QY 601 GluIleLysGluArgTyrPheTyrThrIleGluGlyGluTyrArgValAlaAlaArgAla 620
Db 1801 GAGATAAAGACGCGTATTTCTATACCGCAGGAGGAGAAATACAGATAGATGCAAGGCT 1860
QY 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
Db 1861 TCTGAGACCATGAGGAACCTCGTACTTTTACAAGATGCTCTACAAGATTTCCCAACATTA 1920
QY 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal 660
Db 1921 TTCAATGGTGGCCAGCCCACTGACAGAGTGGTCAACAAATGATCACACCATTAGACGTC 1980
QY 661 ProProLeuAspTyrPhe 666
Db 1981 CCACCATTAGACTACTTC 1998

```

RESULT 7
CR380947_0

WPCOMMENT
Sequence split into 5 fragments LOCUS CR380947 Accession CR380947
Fragment Name Begin End
CR380947_0 1 110000
CR380947_1 100001 210000
CR380947_2 200001 310000
CR380947_3 300001 410000
CR380947_4 400001 485192
LOCUS CR380947 485192 bp DNA linear PLN 30-JUL-2004
DEFINITION Candida glabrata strain CBS138 chromosome A complete sequence.
ACCESSION CR380947
VERSION CR380947.1 GI:49524079
KEYWORDS
SOURCE
ORGANISM
Candida glabrata CBS138
Candida glabrata CBS138
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS
Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S.,
Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E.,
Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A.,
Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Bayne, E.,
Blekasten, C., Boisarne, A., Boyer, J., Cattolico, L.,
Confanier, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C.,
Ferry-Dumazet, H., Groppi, A., Hantre, F., Hennequin, C.,
Jaubert, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R.,
Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.M., Nikolai, M.,
Ozta, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potter, S.,
Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekala, F.,
Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M.,
Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C.,
Caudron, B., Scarpetti, C., Gaillardin, C., Weissenbach, J., Wincker, P.,
and Souciet, J.L.
Genolevures
Genome evolution in yeasts
Nature 430 (6995), 35-44 (2004)
15229592
2
Frangeul, L. and Sherman, D.
Direct Submission
Submitted (05-MAY-2004) B. Dujon, Institut Pasteur, Unite de
Genetique Moleculaire des Levures, 25 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: lfrangeul@pasteur.fr E-mail:
david@labri.fr
This is the first release of the complete genome sequence. The
Candida glabrata sequence is made of 13 chromosomes (A to M). These
sequences contains 6 gaps (artificially closed) indicated by 'gap'
features and about ten colinearity problems indicated by 'gap'
'misc difference' features. An update of the sequence will be
submitted in few weeks. For more information go to
http://cbi.labri.fr/Genolevures.
Location/Qualifiers
1. 485192
/organism="Candida glabrata CBS138"
/mol_type="genomic DNA"
/strain="CBS138"
/db_xref="taxon:284593"
/chromosome="A"
complement(1608..2765)
/locus_tag="CAGLOA00110g"
/pseudo
complement(1608..2765)
/locus_tag="CAGLOA00110g"
/note="some similarities with sp|P08640 Saccharomyces
cerevisiae YK019c STAl, no start, frameshift"
/pseudo
complement(2671..4809)
/locus_tag="CAGLOA00121g"
complement(2671..4809)
/locus_tag="CAGLOA00121g"
/note="some similarities with sp|P38894 Saccharomyces


```
QY 43 IIEHISGLUPHEASPPTRTPHEASNTYRARGALATHRYLSTYRLEUVALASNSER 62
Db 22392 ATTATGATATGACCCCTGGTTCAACTTCAGGCTCAAAATAACCTGGTTTCCACATTCG 22451
QY 63 PHEIYRLEUASNTYRPHASPPTRTPHEASNTYRTPYRPROLEUGLYARGVALTHR 82
Db 22452 TTCCACGAGTCTCTGAACTGGTCGACGATAGACATGGTATCAATTTGGGTGAGTCACT 22511
QY 83 GLYGLYTHRLTYRPROGLYLEUMETTHRTHRSERLAPHELETHRPHISALALEUARG 102
Db 22512 GGTGTACTTTGTACCCCTGGTTGATGACCAAGTGTCTTTGTATGGCCACATCGTGGCCG 22571
QY 103 ASNTYRLEUGLYLEUPROILEASPILEARGASNVALCYSVALLEUPHEALAPROLEUPHE 122
Db 22572 AAG---ATCGGTCTACCTGTTGACATTAGAAACATCTGTGTCTTTTGTCTGCTGCTT 22628
QY 123 SERGLYVALTHRALATHRPHALATHRGLUPEHETHRGLULILEYASAPALASERALA 142
Db 22629 GCCGGTATTACCGCTATCGCACTTACCAATTCACAAAGGAATGGAAGGATTCCTCCGCC 22688
QY 143 GLYLEULEUALAAGLYPHEILEALALEVALPROGLYTYRILESERARGSERVALALA 162
Db 22689 GGTCTACTGGCTGCTGTCTCATGCTATTCTCCAGGTTACATTTCCAGATCTGTGCT 22748
QY 163 GLYSERYRASPENGLUALALEALALETHRLTYRLEUETVALTHRPHETPHE 182
Db 22749 GGTCTCTACGATACGAAGCCATCTCAATTAATCTATGATGATAACCTTCCTGCTGG 22808
QY 183 ILELYSALAGINLYSTHRYLSEYRILEMETHISALATHRCYSALALEUPHEIYRPH 202
Db 22809 ATCAAAGCACAAAGACTGGGTCCATCATGCAAGCCACTTTGCTGCTGCTTCTACTTT 22868
QY 203 TYRMETVALSERALATTPGLYTYRVALPHEILETHRASNLEULILEPROLEUHSVAL 222
Db 22869 TACATGGTCTCAGCTGGGGTGGTACGCTCTTACTTACTTAAATTCATTTACACGTT 22928
QY 223 PHELEULEULEUETGLYARGTYRSEYRILEYRLEUVALTYRTHRTRTP 242
Db 22929 TTATCTCTATCTAATGGCGGTATATAACTCTAGATTATACATGCTTACTCCACTGG 22988
QY 243 TYRALALEGLYTHRVALALASERMETGLINLEPROHEVALGLYPHELEUPROILEARG 262
Db 22989 TACGCAATGGTACATGGCTCTATGCAAAATCCCTTTGGGGTTTCTTGGCCAAATCAGA 23048
QY 263 SERASNAPHSIMETALALEUGLYVALPHEGLYLEULEGGLINILEVALALAPHEGLY 282
Db 23049 TCTATGACCATATGGCTGCTGGTGTCTTGGCTTCTATTCDAATGTTGCTTGGT 23108
QY 283 ASPHEVALLYSGLYGLINILESERTHRALALYPHELYSVALILEMETVALSERLEU 302
Db 23109 GACTTCGTCAGGCGCAGAGTCCCGCACAAAAGTTCAGAGTCACTCTCTTGGTTTCTTTA 23168
QY 303 PHELEULEULEUVALGLYVALGLYLEUSERALALEUHTYRMETGLYLEULE 322
Db 23169 TTGGTGTGATGGCTGATGATTTTGGGTTTGTGCTTGTGCTTACTTCTGTTGAT 23228
QY 323 ALAPROTHRGLYARGPHEIYRSEYRILEYRLEUVALTYRASNLYRLEUHSILE 342
Db 23229 GCCCATGACCTGGTAGATTTTACTCTCTTGGGATACGCTACGCTAAGGTTCACT 23288
QY 343 PROILEILEASERVALSERGLUHSINPROVALSERTPRPROALAPHEPHEASP 362
Db 23289 CCTATCAATTCCTCTGCTCCGACATCAACAGTAGCTGGCCATCTTCTTTTGTGAC 23348
QY 363 THRHSIPHELEULEITHRLEUPHEPROALAGLYVALPHELEUETVALSERLEULYS 382
Db 23349 CAACACTTTTGTATCTGGTGTTCACAGCGGTGTTTCTTGTGTTTCAATGGAATGAAG 23408
QY 383 ASPGLUHSIPHEVALPHEVALILEALATYRSEYRILEYRLEUVALGLYVALMET 402
Db 23409 GATGACGAGCTTTGTTATCTGTATTCTGTTTCACTTCTGTTTCACTGCTGTTATG 23468
QY 403 VALARGLEUETLEUTHRTHRPROVALILECYSVALSERALALEUVALALEUSER 422
```

RESULT 8

CR382123_09/c

WPCOMMENT

Sequence split into 18 fragments LOCUS CR382123 Accession CR382123
Fragment Name Begin End
CR382123_00 1 110000

Qy 26 GlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGluSerIleHisGlu 45
Db 112532 GAGGGGCAATCTCGTACGACTGTTTCGGTAAATCCGGTTCGAGTCCATCCACGAG 112591
Qy 46 PheAspProTrpPheAsnTrpArgAlaThrLysTrpLeuValAsnAsnSerPheTrpLys 65
Db 112592 TTCGACCCCTGGTGTAACTTCGCGCACACACAGTACCTGTGTGACCCACTCGTTCACGAG 112651
Qy 66 PheLeuAsnTrpPheAspAspArgTrpTrpTrpProLeuGlyArgValThrGlyGlyThr 85
Db 112652 TTTCGAAATGTTGACGACCGGACGTTGGTACCGCTAGACCGGTGAGGGGGGAGC 112711
Qy 86 LeuTrpProGlyLeuMetThrThsSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeu 105
Db 112712 CTGTACCCGGGCTCATGACCACTTCCTGCGCTGATCTGGCACCGGTTCGGAAGTTC--- 112768
Qy 106 GlyLeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyVal 125
Db 112769 GGGCTGCCAGTGGACATCCGGAACATCTGTGTCTGTTTCGCGCCGCTTGGCGGGCTG 112828
Qy 126 ThrAlaTrpAlaThrTyGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeu 145
Db 112829 AGCGCTGCGGACGTATGAGTTTACGAGGAATCAAGGACGAGAGCGGGCTGCTG 112888
Qy 146 AlaAlaGlyPheIleAlaIleValProGlyTrpIleSerArgSerValAlaGlySerTrp 165
Db 112889 GCTGCGGGCTTCATGCGGATAGTCCCGGTATCATCTCGCGGTCTGTGGCGGCTCTTAC 112948
Qy 166 AspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAla 185
Db 112949 GACACGAGGCGATCGCATCACGCTGCTGATGTACGTTCAGTTCGAGTCAAGCGC 113008
Qy 186 GlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPheTrpPheTrpMetVal 205
Db 113009 ATCAAAACGGGCTCGATACCTCACTCCACGCTAGCGCGCTCTTCTACTTCTACATGTT 113068
Qy 206 SerAlaTrpGlyGlyValPheIleThrAsnLeuIleProLeuHisValPheLeuLeu 225
Db 113069 TCTGCTGGGAGGCTAGCTTTTATCACGAACTGATCCGCTGACGCTGTTCATCTG 113128
Qy 226 IleLeuMetGlyArgTrpSerSerLysLeuTrpSerAlaTrpThrTrpTrpAlaIle 245
Db 113129 ATTCTGATGGCGGCTCAAGAGTAAGCTGTACAGCGGTACACTACGTTGATGCGATC 113188
Qy 246 GlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAsp 265
Db 113189 GGGACTGTGGCTCCATCCAGATACCGTTTGTGGGTTTTCGGATTGCTCCATGAC 113248
Qy 266 HisMetAlaAlaLeuGlyValPheGlyLeuGlnIleValAlaPheGlyAspPheVal 285
Db 113249 CACATGGCTGGCTTCGGGCTTTTGGGCTCATTCAGTGTGCTGGGTGGGACACTACATG 113308
Qy 286 LysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeu 305
Db 113309 AAGTCCCATGTCTCCACCGATAAGTCAAGATCGCACTCCATCTCCGCGCGGTGTT 113368
Qy 306 LeuValLeuGlyValValGlyLeuSerAlaLeuThrTrpMetGlyLeuLeuAlaProTrp 325
Db 113369 GGGGGGATGGAGTGTGGGCTGTGTGCGCTGACCATGTGTGGGTATCGCTCTTGG 113428
Qy 326 ThrGlyArgPheTrpSerLeuTrpAspThrAsnTrpAlaLysIleHisIleProIleLeu 345
Db 113429 ACCGGCGGTTTTACTCGCTTTGGACACAACTACGGAAGATCCACATCCCATTCAT 113488
Qy 346 AlaSerValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPhe 365
Db 113489 GCCTCGGTGTCGAACACACGCTACGCGGTGGCTGCTCATTTTCTTGACATACATGTT 113548
Qy 366 LeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHis 385
Db 113549 CTCATCTGGCTCTTCCCGCGGGCTCTTCTCTATTTTGGACCTGAAGGACGAGCAC 113608

Qy 386 ValPheValIleAlaTrpSerValLeuCysserTrpPheAlaGlyValMetValArgLeu 405
Db 113609 GTCTTCGTGATCGTGTACTCGTGTGTGTCTGTACTTCGCGGCGGTGATGGTGTCTG 113668
Qy 406 MetLeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePhe 425
Db 113669 ATGCTGACCTGACCCCATCATCTGTGTGCGCGGATGATGGCTCTCCAAGCTCTTC 113728
Qy 426 AspIleTrpLeuAspPheLys-----ThrSerAspArgLysTrpAlaIle 440
Db 113729 GACGAGTACTCGACTTCTCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 113788
Qy 441 LysProAlaAla-----LeuLeuAlaLysLeuIleValSerGlySer 454
Db 113789 CGCCCTTCGCGCGCGGAGGGTTCCTGAACTGTCTGGAAGCTCTCTGATCAGGCACC 113848
Qy 455 PheIlePheTrpLeuTrpLeuPheValPheHisSerThrTrpValThrArgThrAlaTrp 474
Db 113849 TTCCTCTTCTACTCTCTCTCTTGTCTACCATCTGACCTGCGGGTCACTCGACCGGCTAC 113908
Qy 475 SerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAsp 494
Db 113909 TCTCGCCCTTCGCTGCTGCTGCTTCGCACTCCGACGGCTCCCGCGCTCATGTAC 113968
Qy 495 AspPheArgGluAlaTrpTrpLeuArgMetAsnSerAspGluAspSerLysValAla 514
Db 113969 GACTACCGGAGGCTTACTTGTGCTCGCATGACACCGACCGCGCGCGCGCGCGCGCG 114028
Qy 515 AlaTrpTrpAspTrpGlyTrpGlnIleGlyMetAlaAspArgThrThrLeuValAsp 534
Db 114029 GCTGTGTGGACTAGGGTACCATCGGCGGCTGCGCGGCGCGCGCGCGCGCGCGCG 114088
Qy 535 AsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu 554
Db 114089 AACACACCTGGAACACACACCATCGCCACTCTCGGCAAGGCCATGGCTCGCGCCAG 114148
Qy 555 GluLysSerTrpGluIleLeuLysGluHisAspValAspTrpValLeuValIlePheGly 574
Db 114149 GACAGGGCTACGAGATCCTTCGCGCGCCACGACCTCGACTACGCTCTCTCTCTCTCG 114208
Qy 575 GlyLeuIleGlyPheGlyValAspAspIleAsnLysPheLeuTrpMetIleArgIleSer 594
Db 114209 GCGTACTCGGCTTCAGCGCGCGAGACCTCAACAACTCTCTGATGGTTCGCGCATCGCA 114268
Qy 595 GluGlyIleTrpProGluGluIleLysGluArgTrpPheTrpThrAlaGluGlyGluTrp 614
Db 114269 GAGGGCATCTGCGCTGACGAGATCCGCGAGCGCAACTACTTCAGCGCGCGCGCGCGAGTAC 114328
Qy 615 ArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTrpLysMetSerTrp 634
Db 114329 CGATGGATGCGCGCGCTCCCGCCCATGAAGAACCGCTCATCTACAAGCTCTCTCTAC 114388
Qy 635 LysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMet 654
Db 114389 CACCGCTTCCCGAGCTCTTCGCGCACACCCGAGCGCGCGCGCGCGCGCGCGCGCG 114448
Qy 655 IleThrProLeuAspValProProLeuAspTrpPheAspGlnValPheThrSerGluAsn 674
Db 114449 ATCTCGCGCGCGCATCGGCTCGCTCGACTACTTCGAGAGGCTTTCACCTCCGAGAAC 114508
Qy 675 TrpMetValArgIleTrpGlnLeuLysAspAspAlaGlnGlyArgThrLeuArgAsp 694
Db 114509 TGGATGTCGGCATCTACAAGCTCAACAGCCGACACCGCTCGCGCGCGCGCGCGCGCG 114568
Qy 695 ValGlyGluLeuThrArg-----SerSerThrLysThrArgArgSerIleLysArgPro 712
Db 114569 GCTAGCGGCTTCGACCGGAGCTCGCGCGCGCGCTCAAGAGAGGCTCTCTCAAGAGCCC 114628
Qy 713 GluLeuGlyLeu 716
Db 114629 GCGTGGGCGCTC 114640

RESULT 11

| | | |
|------|--|------|
| 1726 | GTTCGTAAGCAATGCTTCCCTGAAGANGTGTGCTATGAATAATTTTGAGACAACACGAT | 1785 |
| 566 | ValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAspIleAsn | 585 |
| 1786 | GTTCATTATGTTAGTTATATTGGAGGGTTATTGGGTTATTCTGGTGATGATATTAAAC | 1845 |
| 586 | LysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArg | 605 |
| 1846 | AAATTCATTATGGATGCTAAGAAATGCTGAAGGTAICTGGCTGATGAATCAAGAAAGA | 1905 |
| 606 | TyrPheTyrThrAlaGluGlyGlnTyrArgValAspAlaArgAlaSerGluThrMetArg | 625 |
| 1906 | GACTACTTTTACTGACCGAGAGAAATAAAGTGGTAAGAATGCATCCTGGCCAAATGAAG | 1965 |
| 626 | AsnSerLeuLeuTrpLysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGln | 645 |
| 1966 | AAATCTTTGATGATAAGTTATGCTATCATAGATTCACTGAATTTGTTGGAGGTAGAGAT | 2025 |
| 646 | AlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyr | 665 |
| 2026 | GGTGTTCATAGATTAGAAACCAACAATCCAGCCCAATGAAGTACCGAAATGAATGTT | 2085 |
| 666 | PheAspGluValPheThrSerGlnAsnTrpMetValArgIleTyrGlnLeuLysLysAsp | 685 |
| 2086 | GTGGAGAGACCTTCATCAGAAATTTGGATTTGAGAAATTTCAAAAGTTTAAAGATTG | 2145 |
| 686 | AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer | 702 |
| 2146 | GATAAATCTGGTATAGATTATCATCAAGCTACTGCTTTTGAAGAAATCATCATCGGCACT | 2205 |
| 703 | ThrLysThrArgSerIleLysArgProGluLeuGlyLeuArg | 717 |
| 2206 | TCCAAAAGAACAGATCCCTAAAGAGACCTTAATTTGGAGTAAGA | 2250 |

SECRET

RESULT 12
CD307134 03

CR382134

WPCOMMENT

| Fragment Name | Begin | End |
|---------------|----------|---------|
| CR382134_00 | 1 | 110000 |
| CR382134_01 | 100001 | 210000 |
| CR382134_02 | 200001 | 310000 |
| CR382134_03 | 300001 | 410000 |
| CR382134_04 | 400001 | 510000 |
| CR382134_05 | 500001 | 610000 |
| CR382134_06 | 600001 | 710000 |
| CR382134_07 | 700001 | 810000 |
| CR382134_08 | 800001 | 910000 |
| CR382134_09 | 900001 | 1010000 |
| CR382134_10 | 10000001 | 1110000 |
| CR382134_11 | 11000001 | 1210000 |
| CR382134_12 | 1200001 | 1310000 |
| CR382134_13 | 1300001 | 1349996 |

Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vansireels, E., Rieger, M., Schaefer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehmann, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purrelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gallardin, C., Talleda, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, J., Jimenez, J. L., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L., Lowe, T., McCombie, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrall, B. G. and Nurse, P.

The genome sequence of *Schizosaccharomyces pombe*
 Nature 415 (6874), 871-880 (2002).

21848401
 11859360
 2 (bases 1 to 40907)
 Lyne, M., Rajandream, M. A., Barrall, B. G., Beck, A., Borzym, K., Klages, S., Langer, I. and Reinhardt, R.

Direct Submission
 Submitted (23-NOV-1998) European Schizosaccharomyces genome sequencing project, Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk and Max Planck Institut fuer Molekulare Genetik, Innestr. 73, D-14195 Berlin, Germany

On Feb 7, 2000 this sequence version replaced gi:3925769.

Notes:
 Details of S. pombe sequencing at the Sanger Institute are available on the World Wide Web
 http://www.sanger.ac.uk/projects/s_pombe/
 CDS are numbered using the following system eg SPAC10.01c. SP (S. pombe), A (chromosome 1), c510 (cosmid name), .01 (first CDS), c (complementary strand). However, clones may have been reorientated since the original submission, therefore the complementary strand notation may be invalid for strand inference. IMPORANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once.

COMMENT

FEATURES
source

1..40907
 /organism="Schizosaccharomyces pombe"
 /mol_type="genomic DNA"
 /strain="972h."
 /db_xref="taxon:4896"
 /chromosome="II"
 /map="11L"
 /clone="cosmid c1271"
 1307..3367
 /gene="SPBC1271.15c"
 1307..3367
 /gene="SPBC1271.15c"
 /note="translation initiation factor if-2 (predicted); localization mitochondrial (predicted); non-essential (PMID 12618370); similar to S. cerevisiae YOL023W"
 /codon_start=1
 /protein_id="CAA22205.1"
 /db_xref="GI:3925784"
 /db_xref="GOA:O59683"
 /db_xref="Swiss-Prot:O59683"
 /translation="MAFLSACTSRVHVNAGFTFTKFPVRNVHKLVSYHSPLSRRF SLADISVKNHNGPQLRLOTHPPHSSPMHATKGFREPKSTPLTFLAVLSASFA NLLQIPTRILRDLNRIGDKVDFEYLITTEYSLAEYGFQDVAAYSPSPKSL NSKKDGLRPVPTLMGVHDHGTLLDFAFRKSTASTEHGKITQKIGAFVFDKX SKFTFLDTPGHAFEMARKGANIADIVLVVAGDGVKPTQVEAIKHQIADVPV VALTKSRPGPFIHKIEQLNNGIOVEALGGETQIIPISAKTGKGIPELEAIIITLA BIMEIRASPPFPQGVIVSVSSVTGKVSATVVVVRGTVKGMVLVAGSWCKVRSVL DWKSKISQVGOAAQVYKQDLPVIAQDLAYEVSEAKRILSDIYRQSEONQFYE LAESCEORVSALAKKSGPAAIQEETSVSNSIIAKCDTGSMEALSDYKLPQFG KYKSRVLYTGVPVTDIERASTSDAILISFGVSPKATFLAEKHNVKLLFHVIV HLMDDVKLFLPPIINVRVYGEAIIISAFIDIKARVAVHVGCRVNTGTEIKSHK IRLVRNDKLIWSGEIDSKLKLBVEVTSIKKRGCGILLNFDIVTGDKLQTVBEYK

gene

CDS

misc_feature

misc_feature

misc_feature

gene

CDS

misc_feature

misc_RNA

gene

CDS

PPDF"
 1811..2815
 /gene="SPBC1271.15c"
 /note="pfam match to entry PF00009 GTP_EFTU, Elongation factor Tu family"
 1814..2815
 /gene="SPBC1271.15c"
 /note="Match to GTP_EFTU, Score 114.20"
 3215..3283
 /gene="SPBC1271.15c"
 /note="PS01176 Initiation factor 2 signature"
 complement(3720..5057)
 /gene="SPBC1271.14"
 complement(3720..5057)
 /gene="SPBC1271.14"
 /EC_number="2.3.1.35"
 /note="amino-acid N-acetyltransferase (predicted); glutamate N-acetyltransferase (predicted); involved in ornithine biosynthesis (predicted); involved in arginine biosynthesis (predicted); amino acid N-acetyltransferase activity; glutamate N-acetyltransferase activity; localization mitochondrial matrix (predicted); non-essential (PMID 12618370); similar to S. cerevisiae YMR062C
 PATH:map00220
 aspect=P; term=ornithine biosynthesis; GOid=GO:0006592; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=P; term=arginine biosynthesis; GOid=GO:0006526; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=F; term=amino acid N-acetyltransferase; GOid=GO:0004042; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=F; term=glutamate N-acetyltransferase; GOid=GO:0004358; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 aspect=C; term=mitochondrial matrix; GOid=GO:0005759; evidence=ISS; db_xref=GO:unpublished; with=SPTR:Q04728; date=20030616
 /codon_start=1
 /protein_id="CAA22204.1"
 /db_xref="GI:3925783"
 /db_xref="GOA:O94346"
 /db_xref="TrEMBL:O94346"
 /translation="MQTKIFARLAGIGRAFVGTMPDKTKFISKGTYPQGSFLNGIA SGVANGKDLALFSEFCNAAAFTRKAFQAPQVSRQINGCGKDHCYVFN SGCANAVTGEGLMDAQLITAEADNLTRPHWTENSEEFPSSLVNMTVIGURLKD KIOSGLEHVEDLGSTHEYMMRAEAICTDTPFKLVSRSLAGKVIAGFAKAG MINPNLATLLGLEFVDAPISVDAVRSILRHAINNSFNISIDGDTSTNDTIAFLANGA AGSEITKSPAYKEIRDAVDTAQQLAKLVWDGEGATKFTVTVRGARSEKDAALV ASTISNALKVTAFFGEDANWGRILCAVYSGAAVNPATTVSFIPADTEPLKLVN GEQNVDCTPSSILSQDELTVDDVDJGCGPYAKTNWTCDFSVDYVRINADYS"
 complement(3723..4946)
 /gene="SPBC1271.14"
 /note="Match to Argu, Score 840.54"
 complement(6265..6931)
 /note="SPMEIRNA D31852 Fission yeast DNA for meiRNA, complete sequence"
 complement(join(7889..8438,8502..8575))
 /gene="SPBC1271.13"
 complement(join(7889..8438,8502..8575))
 /gene="SPBC1271.13"
 /note="mitochondrial ribosomal protein L17; essential (PMID 12618370); similar to S. cerevisiae YJL063C"
 /protein_id="CAA22203.1"
 /db_xref="GI:3925782"
 /db_xref="GOA:O94345"
 /db_xref="TrEMBL:O94345"
 /translation="MTKTYIRKLGRSTSAHQRLRLNLTSLVKHESIQTTWAKAEA ORFAEKLIITWAKRANPQNNRKLAEQVFEKSTLLKKVDFVLVPRYNGRCGYTRLLK

LPRSTDNAPMGVLEFVDGPKDIRHMTAKTVGICLQAKALPTIRKNIHKVLLFRK
 NGXARFDQLVQKESSEHARLKHEDHEKTVKDWKRDPIPRPTIYI
 /gene="SPBC1271.13"
 /notes="Match to Ribosomal L17, Score 114.95"
 /notes="SPBC1271.13"
 /gene="SPBC1271.13"
 /notes="cctaataaagggaacag, splice branch and acceptor"
 /complement (8496, 8501)
 /gene="SPBC1271.13"
 /notes="Gtatgc, splice donor sequence"
 /complement (8899, 9327)
 /notes="mRNA from RH016222"
 /complement (9328, 9869, 9925, 10241, 10424, 10546,
 10690, 10768, 10941, 10966, 11042, 11054, 11105, 11175)
 /gene="keel"
 /notes="synonym: SPBC1271.12"
 /complement (9328, 9869, 9925, 10241, 10424, 10546,
 10690, 10768, 10941, 10966, 11042, 11054, 11105, 11175)
 /gene="keel"
 /notes="oxysterol binding protein (predicted); involved in
 ergosterol biosynthesis (predicted); similar to S.
 cerevisiae YHR001W and YHR003W and YOR237W and YPL145C;
 similar to S. pombe SPBC646.08c and SPBC354.07c and
 SPCC2386.01; localization golgi (predicted); non-essential
 (PMID 12618370)"
 /codon_start=1
 /protein_id="CAA22202.1"
 /db_xref="GI:3925781"
 /db_xref="GOA:O74178"
 /db_xref="TrEMBL:O74178"

Alignment Scores:

Pred. No.: 9,646-176 Length: 40907
 Score: 2237.00 Matches: 425
 Percent Similarity: 72.84% Conservative: 114
 Best Local Similarity: 57.43% Mismatches: 167
 Query Match: 59.48% Indels: 34
 DB: 8 Gaps: 5

US-10-028-384-6 (1-718) x SPBC1271 (1-40907)

QY 7 CysValLeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGly 26
 DB 37681 TGCATAATAGCTGTATTAGCATTGGTAAATCAAGTATCTTTAGT----- 37634
 QY 27 AlaAlaIleSerArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPhe 46
 DB 37633 GCTTCGTCTCTCGAGGCTTTTCTGTCATTCTGTTACGAGTCCATTCATGAAATT 37574
 QY 47 AspProThrPheAsnTyrAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPhe 66
 DB 37573 GATCCTTGGTCAATTCGAGCTCCAAATATTGGTGGAAACAAAGGTTTATAACTTT 37514
 QY 67 LeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArgValThrGlyClyThrLeu 86
 DB 37513 TTAATTTGGTTTCATGAAGAAGTTGTTACCCGTTGGTCGTTAGCGGGTGTACTTTG 37454
 QY 87 TyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGly 106
 DB 37453 TACCAGAGCTTATGGTCAGCTGTGATTATTTTCAAAAGTTTAA---CATCTTTTAA 37397
 QY 107 LeuProIleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThr 126
 DB 37396 ATTAACGTGAACATTCTGTATGATGTATGTTTACTTGGCCCTGCTTCTCTGGAATCACT 37337
 QY 127 AlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuAla 146
 DB 37336 CGGATTCTACCTATTATCTGGCTAGAGAATTGAAAGATGATGATGCGCTTTTAGCT 37277
 QY 147 AlaGlyPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAsp 166
 DB 37276 GCCGCAITTTATGGGTATTTCTCTCGATACACCTCCGTTCCGTCGTTCTTACGAT 37217

QY 167 AsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheThrIleLysAlaGln 186
 DB 37216 AATGAAGCAATTCCTATTACCTTTTGATGCAACAGTTGCTTTGGGATCAAGGAGTG 37157
 QY 187 LysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSer 206
 DB 37156 AAGTCTGGCTCTCTTTCTGGGGTGGCTGCACAGATTGCTACTTCTATATGTTAACT 37097
 QY 207 AlaTrpGlyGlyTyrValPheIleThrAsnLeuLeuProLeuHisValPheLeuLeu 226
 DB 37096 GCGTGGGGTGGTATGATTATCATCAAAATGATACCTTTACAGTATTGTTCTCTA 37037
 QY 227 LeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGly 246
 DB 37036 CTTATGGTGGCTATCTACTAGCAATATATACATTCTTACACACATATGTTATTGGA 36977
 QY 247 ThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleAsgSerAsnAspHis 266
 DB 36976 AGCTGGCTCTATCAAGTTCGGTTTGGTTTCCACCGGTGCGATGATGACAT 36917
 QY 267 MetAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLys 286
 DB 36916 ATGTCGGTTTAGGAGTCTTTGGCTGTACAGCTTTTTCATTTCTACAATATGTTAAA 36857
 QY 287 GlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuLeuLeu 306
 DB 36856 GGTCTAGTTTATCCCAAGCAATTCAAATCTATTGTTGTTGCTTGGTTCTTAGTG 36797
 QY 307 ValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThr 326
 DB 36796 GGTCTAGCAACAGTCTCTTTTCTTATCTTCAACAGGTGTATCGCTCTCTTGACA 36737
 QY 327 GlyArgPheThrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleAla 346
 DB 36736 GAGCGTTTCTATCTCTTTGGATACAACTACCCCAAGATTCAATCTCTATCATGCT 36677
 QY 347 SerValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeu 366
 DB 36676 TCGGTATCAGAACATCATCGCTCTCTTCTTCAACAGGTGTATCGCTCTCTTGACA 36617
 QY 367 IleTrpLeuPheProAlaGlyValPheLeuLeuLeuAspLeuLysAspGluHisVal 386
 DB 36616 ATTGGTTATTGCCAGTTGGTGTACTTGTGTTTCAAGGACCTCGTAAGAACATGTC 36557
 QY 387 PheValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMet 406
 DB 36556 TTTATTATATATATCTCTTAGCAACATATTTTGTGGTGTGATGGTTCGTTGGTT 36497
 QY 407 LeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAsp 426
 DB 36496 TTAACCTTAACCTCTGTTGTGATAGCTCTGCTGTAGCAATTCCTCTTTTAGAC 36437
 QY 427 IleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIleLysProAla----- 443
 DB 36436 ACATATATGGTCTCTGAAGTTGAAGAGGACAAAGTGACGGAAGAGCGCTTCAGCCAAA 36377
 QY 443 ----- 443
 DB 36376 TCTAAGAACAGAAAGGTATTTCTCTATTCTTATGTTTCTTCTCTGCTCAAAAAAT 36317
 QY 444 -----AlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeu 459
 DB 36316 ATTGGAATTTACAGTTTCCTTTCCAGAGTATTAGTCAATTCCTCTACCGCATATTCCTA 36257
 QY 460 TyrLeuPheValPheHisThrTrpValThrArgThrAlaTyrSerSerProSerVal 479
 DB 36256 ATAAGTTTGTATTATCTCCAGTTGGGTGACTCTAATGTTTACTTCTCCCTACCGTG 36197
 QY 480 ValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspPheArgGluAla 499
 DB 36196 GTTTTGTCTACCGTGTAAACGATGAGTAGTTAATGATATATATGATGACTTCCGTAAGCT 36137

| | | | |
|------|----|---|------|
| 1721 | Db | CGCCGATTTATGGGATATGCTCTCGATACACTCCCGTTCCGTCGCTGGTTCCTAGCAT | 1780 |
| 167 | QY | AsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGln | 196 |
| 1781 | Db | AATGAAGCAATGTGCTATATCCCTTTTCATGTCGAAAGCTTGTCTTGTGATCAAGCGAGT | 1840 |
| 187 | QY | LysThrGlySerIleMetHisA1aThrCysAlaAlaLeuPheTyrPheTyrMetValSer | 206 |
| 1841 | Db | AGATCGGCTCCCTTTCTGGGGTGCCTGCACAGGATTCCTACTCTATATGGTAAT | 1500 |
| 207 | QY | AlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeu | 226 |
| 1901 | Db | CGCGGGGTGGTATGATATTATCATCAAAACATGATACCTTTACAGTATTTGTTCCTTA | 1960 |
| 227 | QY | LeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGly | 246 |
| 1961 | Db | CTTATGGTGCCTATACTAGCAAAATATATACATGTGCTACACACATACTATGTATTGGGA | 2020 |
| 247 | QY | ThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleA1aGlySerAsnAspHis | 266 |
| 2021 | Db | ACGCTGGCTTCATGCGANGTCCGTTGTGGTTCCACCCGGTGCAGCTAGTAGCAT | 2080 |
| 267 | QY | MetalAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLys | 286 |
| 2081 | Db | ATGCCCCGCTTTAGAGGTGTTGGCCCTTTACAGCTTTTTCATCTCTCAATATATGTTAAA | 2140 |
| 287 | QY | GlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuLeu | 306 |
| 2141 | Db | GGTCTAGTTTCATCCAAAGCAATTCCAAATACTATTTCGTTTTCGCTGGTTGCTTAGTG | 2200 |
| 307 | QY | ValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuAlaProTyrThr | 326 |
| 2201 | Db | GGTCTAGCAACAGTCGCTTTTGTCTTATCTTCAACAGGTGTATCGCTCTTGGACA | 2260 |
| 327 | QY | GlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAla | 346 |
| 2261 | Db | GGAGCTTTCTATCTCTTTGGGATACAAACTACGCCAAGATTCTATATTCCTCATTCGT | 2320 |
| 347 | QY | SerValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeu | 366 |
| 2321 | Db | TCGGTATCAGAAACATCAGCCTCCTACTCGAGTTCGTTCTTCTTGAATTCATATTTTG | 2380 |
| 367 | QY | IleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisVal | 386 |
| 2381 | Db | ATTTCGTTATTCGCCAGTGGTGTTACTTGTGTTTCAAGGAACCTCGTATAGCAATGTC | 2440 |
| 387 | QY | PheValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMet | 406 |
| 2441 | Db | TTTATTATTATATATCCTGCTCTAGAAACATATTTTGTGGTGTGATGGTTTCGTTGGTT | 2500 |
| 407 | QY | LeuThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAsp | 426 |
| 2501 | Db | TTAACCTTACTCCTTGTGTTTCATAGCTGCTGCTGTAGCAATTTCCACTCTTTTAGAC | 2560 |
| 427 | QY | IleTyrLeuAspPheLysThrSerAspA1aGlyTyrAlaIleLysProAla | 443 |
| 2561 | Db | ACATATATGGGTCTCGAAGTTGAAGAGCAAGATGAGCGAAGAGCGCTTCAGCCAAA | 2620 |
| 443 | QY | ----- | 443 |
| 2621 | Db | TCTAAGAACAAAGATATTCTCTATTTCTTAGTTTCTTCACTCTGGCTCAAAAAT | 2680 |
| 444 | QY | -----AlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeu | 459 |
| 2681 | Db | ATTGGATTATACAGTTTCTTCCAGATATTAGTCATTTCTCTACCGCATATTTCCTTA | 2740 |
| 460 | QY | TyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerVal | 479 |
| 2741 | Db | ATAATATGTTGTTATCAATCCAGTGGGACTCTTAATGCTTACTCTTCCCTACCGTG | 2800 |
| 480 | QY | ValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAla | 499 |
| 2801 | Db | GTTTGTGCTACCGGTGTTAAACGATGCTAGTATTAAATGATATTATGACTTCCGTGAAGCT | 2860 |

Search completed: December 15, 2004, 05:47:27
Job time : 8915.54 secs

| | | | |
|------|----|--|------|
| 500 | QY | TyrTyrTrpLeuArgMetAsnSerSerGluAspSerLysValAlaLaIaTrpTrpAspTyr | 519 |
| 2861 | DB | TATGACTGGCTTCGTGAACAACACTCCTTATGACACAAAGGTTATGACGATTCGGTGGGATTAT | 2920 |
| 520 | QY | GlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsn | 539 |
| 2921 | DB | GGTTACCAAAATTCCTGGTATGCGCTGATCGTATTACTTTAGTCGACACAATACGTGGAAAC | 2980 |
| 540 | QY | AsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGlu | 559 |
| 2991 | DB | AACACACATATTCCCACTGGAAAGCCATGCTCTTCCCTGAGAAAAAGCTTACCCT | 3040 |
| 560 | QY | IleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPhe | 579 |
| 3041 | DB | ATCCTCCGTAAACACGATGTTCATTATATTCCTTATATATGCTGGTACTCTTGGATAC | 3100 |
| 580 | QY | GlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpPro | 599 |
| 3101 | DB | AGCAGCGACGACATGAACAAAGTTCCTTTGGATGATCCGAATTTCTCAGGAGTATGCGCC | 3160 |
| 600 | QY | GluGluIleLysGluArgTyrPheTyrThrAlaGlyGlyGluTyrArgValAspAlaArg | 619 |
| 3161 | DB | GATCAAAATAGTAGAGCGTAACTTTTTTACTCCATAATGAGAAATATCGAACTGACGATCGC | 3220 |
| 620 | QY | AlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGln | 639 |
| 3221 | DB | CGTACTCCCACTATGGTGNGCTTTATTATATAGATGTCATATCACGGTGCTTGGAAA | 3280 |
| 640 | QY | LeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAsp | 659 |
| 3281 | DB | CTTTTCCTCCCAATCAAGGATATGACCGTGCTGCCAATCAAAAACATCACCATCGAAGAT | 3340 |
| 660 | QY | ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIle | 679 |
| 3341 | DB | --CCTCAACTATTACTTCGAAGAGCAITTCACCTACCGTTCATCATTTAGTTCGTTTG | 3397 |
| 680 | QY | TyrGlnLeuLysLysAspAlaGlnArgThrLeuArgAspValGlyGluLeu--- | 699 |
| 3398 | DB | TATAAAGGTAAAGAACCGGATACCTTGACGCGATTGTGAACCAAGTGACATT-ATTGGA | 3456 |
| 699 | QY | ThrArgSerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal | 718 |
| 3457 | DB | ACAAGGCAAAAGAAAGATCGCGCGCTCTCGAAAAACTAACGAAATTCCTTTGAGGGTT | 3516 |

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 ; Search time 827.493 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761

Sequence: 1 MGSDRSCVLSVFQILKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10028384/runat.14122004.131518.8007/app_query.fasta_1.3740
-DB=N_Geneseq_23Sep04 -QPMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384.CCGN.1.1986 @runat.14122004.131518.8007 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_23Sep04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 3761 | 100.0 | 2733 | 10 | Add94787 Yeast STT |
| 2 | 3488 | 92.7 | 2000 | 10 | ACC61293 Gene sequ |
| 3 | 3488 | 92.7 | 2000 | 10 | Adk63659 Disease t |
| 4 | 2738 | 72.8 | 2256 | 6 | Abt20789 Candida a |
| 5 | 2428.5 | 64.6 | 2232 | 8 | Abt20789 Aspergill |
| 6 | 2301.5 | 61.2 | 2455 | 10 | Adb69900 C. neofo |

| | | | | | |
|----|--------|------|------|----|--------------------|
| 7 | 2264.5 | 60.2 | 2603 | 8 | ABT20191 Aspergill |
| 8 | 2264.5 | 60.2 | 3969 | 8 | ABT17781 Aspergill |
| 9 | 2264.5 | 60.2 | 4603 | 8 | ABT19595 Aspergill |
| 10 | 2090 | 55.6 | 1848 | 8 | ABT18969 Aspergill |
| 11 | 2042 | 54.3 | 1369 | 8 | ABT18375 Aspergill |
| 12 | 2032.5 | 54.0 | 2882 | 10 | ADB69539 C. neofo |
| 13 | 2032.5 | 54.0 | 4738 | 10 | ADB69178 C. neofo |
| 14 | 2027.5 | 53.9 | 2417 | 10 | ADD94789 Drosophil |
| 15 | 2027.5 | 53.9 | 2699 | 4 | ABL13247 Drosophil |
| 16 | 2013.5 | 53.5 | 2710 | 10 | ADD94785 Mouse SIM |
| 17 | 2012.5 | 53.5 | 2481 | 10 | ADD94783 Mouse SIM |
| 18 | 1955.5 | 52.0 | 4922 | 4 | ABL13246 Drosophil |
| 19 | 1945 | 51.7 | 3094 | 10 | ADD94791 Mouse ITM |
| 20 | 1920 | 51.1 | 2472 | 10 | ADD94793 Human ITM |
| 21 | 1920 | 51.1 | 2472 | 10 | ADD94793 Human chr |
| 22 | 1920 | 51.1 | 2760 | 5 | ABV24502 Human pro |
| 23 | 1902.5 | 50.6 | 2855 | 4 | ABL02795 Drosophil |
| 24 | 1665 | 44.3 | 6153 | 4 | ABL02794 Drosophil |
| 25 | 1259 | 33.5 | 1371 | 12 | ADP28508 Human sec |
| 26 | 1186.5 | 31.5 | 1664 | 4 | AAK94164 Human ful |
| 27 | 1186.5 | 31.5 | 1664 | 12 | ADL30661 Full leng |
| 28 | 1156 | 30.7 | 1543 | 2 | AAK85055 Human sec |
| 29 | 1156 | 30.7 | 1543 | 8 | ACD18981 Novel hum |
| 30 | 1156 | 30.7 | 1543 | 12 | ADG78372 Human sec |
| 31 | 1156 | 30.7 | 1543 | 12 | ADN60663 Human sec |
| 32 | 1154 | 30.7 | 1209 | 4 | AAH33264 Human col |
| 33 | 1154 | 30.7 | 1209 | 6 | ABL89850 Human pol |
| 34 | 739 | 19.6 | 1114 | 4 | AAH99794 Human pro |
| 35 | 719.5 | 19.1 | 2547 | 4 | AAO8315 Human sec |
| 36 | 719.5 | 19.1 | 2660 | 6 | ABO54750 Human ova |
| 37 | 718.5 | 19.1 | 2510 | 5 | AAf93772 Human CDN |
| 38 | 718.5 | 19.1 | 2537 | 4 | ADN08289 Human sec |
| 39 | 718.5 | 19.1 | 2546 | 2 | AAV44866 Clone C75 |
| 40 | 718.5 | 19.1 | 2546 | 5 | AAf98463 Human CDN |
| 41 | 711.5 | 18.9 | 2284 | 4 | AAH18021 Human CDN |
| 42 | 675 | 17.9 | 381 | 6 | ABN23353 Human ORF |
| 43 | 664 | 17.7 | 787 | 4 | AAH07526 Human CDN |
| 44 | 637.5 | 17.0 | 2953 | 4 | ABL18224 Drosophil |
| 45 | 620 | 16.5 | 764 | 5 | AAf93368 Primer sp |

ALIGNMENTS

RESULT 1

ADD94787

ID ADD94787 standard; DNA; 2733 BP.

XX

AC ADD94787;

XX

DT 29-JAN-2004 (first entry)

XX

DE Yeast STT3 gene sequence.

XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cystostatic; immunosuppressive; antineoplastic therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; Yeast;
KW gene; ds; STT3.

OS Saccharomyces cerevisiae.

XX

FN WO2003054008-A2.

XX

PD 03-JUL-2003.

XX

PF 18-DEC-2002; 2002MO-CA001967.

XX

PR 20-DEC-2001; 2001US-00028384.

XX

PA (COMP-) COMPATIGENE INC.

```

XX PI Perreault C, McBride K;
XX XX
XX WPI; 2003-559122/52.
XX DR P-PSDB; ADD94788.
XX XX
XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
XX PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
XX PT or breast cancer, or for suppressing an immune response in an autoimmune
XX PT disease.
XX PS
XX PS Disclosure; SEQ ID NO 5; 66pp; English.
XX XX
XX CC This invention relates to a novel isolated or purified human protein,
XX CC termed source of immunodominant major histocompatibility complex (MHC) -
XX CC associated peptide (SIMP), which is expressed ubiquitously in human
XX CC cells, where the protein has the potential of generating several protein
XX CC fragments binding with high affinity to a human leukocyte antigen (HLA)
XX CC molecule. The invention may allow development of therapeutics with
XX CC cytostatic or immunosuppressive activity or provide sequences useful for
XX CC antisense therapy or gene therapy. The source of immunodominant MHC-
XX CC associated peptide (SIMP) nucleic acids, proteins and fragments are
XX CC useful for diagnosing and treating cancers, for example lung cancer,
XX CC intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast
XX CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
XX CC proteins are also useful for modulating an immune response. Decreasing
XX CC lymphoid cell proliferation is useful for suppressing an immune response
XX CC responsible for an autoimmune disease or a transplant rejection. The
XX CC present sequence is that of the yeast SPM3 gene which is related to the
XX CC invention. Note: this sequence does not appear in the specification but
XX CC was obtained by the indexer from GenBank.
XX SQ Sequence 2733 BP; 712 A; 597 C; 583 G; 841 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 2733
Score: 3761.00 Matches: 718
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ADD94787 (1-2733)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleuLeuValIle 20
DB 219 ATGGGATCCGACCGGTGGTGTGTTGCTGTGTTTCAGACCATCTCAAGCTCGTCATC 278
QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
DB 279 TTGGTGGCGATTTTGGGGTGGCGATATCATCACGTTTGTTCAGTCATCAAAATTGAG 338
QY 41 SerIleIleHisGluPheAspProTTPPheAsnTyrArgAlaThrIysTyrLeuValAsn 60
DB 339 TCTATTATCCATGAATTCGACCCCTGGTTCAATATATAGGGCTACCAAAATATCTCGTCAAC 398
QY 61 AsnSerPheTyrLysPheLeuAsnTTPPheAspArgThrTTPPProLeuGlyArg 80
DB 399 AATTTCGTTTTACAAAGTTTTTGAAGTGGTTTGAAGCGACCTAGTGGTACCCCTCGGAAG 458
QY 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleThrHisAla 100
DB 459 GTTACTGGAGGACATTTATATCTGTTTGTATGACGACCTAGTGGTTCATCTGGCACGCC 518
QY 101 LeuArgAsnThrLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
DB 519 CTGGCAACTGTTTGGGCTTGGCCATTGACATCAGAAACGTTTGTGTCTATTGGCCCA 578
QY 121 LeuPheSerGlyValThrAlaTTPAlaThrTyrGluPheThrIysGluIleLysAspAla 140
DB 579 CTAATTTCTGGGTCCACCGCTGGCGACTTACGAATTTACGAAAGAGATTAAGATGCC 638
QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160

639 AGCGCTGGCGCTTTGGCTGCTGTTTATAGCCATTGTCCCGGTTATATATCTAGATCA 698
161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
699 GTGGCGGGGTCTTACGATTAATAGGCCATTGGCATTACATTAATAGTGCACITTCATG 758
181 PheTTPPileLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
759 TTTTGATTAAAGGCCCAAAAGACTGGCTCTATCATGCACGCAACGCTGTGCAGCTTTATTC 818
201 TyrPheTyrMetValSerAlaTTPGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
819 TACTTCTACATGGTGTGGCTTGGGTGGATAGTGTTTCCCAAACTGTAATCTGCCTACACC 878
221 HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
879 CATGCTCTTTGCTGATTGTGATGGGCAGATATTCGTCAAACTGTAATCTGCCTACACC 938
241 ThrTTPPAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
939 ACTTGGTACGCTATTGCAACTGTTCATCCATGCAGATCCCATTTGTCGGTTTCTACCT 998
261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
999 ATCAGGTCTAACGACCAACATGGCCGATTTGGGTGTTTTCGGTTGATTGATTCGCGC 1058
281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetVal 300
1059 TTCGGTGACTTCGTGAAGGGCCAAATCAGCAGACTAAGTTTAAAGTCATCATGATGGTT 1118
301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
1119 TCTCTGTTTTGATCTTGTGCTTGTGTCGAGCTTTCGCTTGACCTATATGCGG 1178
321 LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle 340
1179 TTGATGGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACATACGCAAGATC 1238
341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe 360
1239 CACATTCCTATCATTCGCTCCGTTTCGGAACATCAACCCGTTTCGTTGGCCGCTTCCTC 1298
361 PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
1299 TTTGATACCCACATTTTGTATCTGGCTATTCGCCGCGGTGATATTCCTACTATTCTCTGAC 1358
381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
1359 TTGAAGACGAGCAGCAGCTTTTGTATCGCTTACTCCGTTCTGTGTTCTGACTTTGCCGGT 1418
401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
1419 GTTATGTTAGTAGTAGTTGATGCTTTGACACCACTCATCTGTGTGCGCGCGCTCGCA 1478
421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
1479 TTGTCCAAAGATTTTGAATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 1538
441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
1539 AAACCTGGGACCTACCTGGCCAAATGATTGTTTCCGGATCATCTCTTTTATTGTTAT 1598
461 LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal 480
1599 CTTTTCGTTCTCCATTTACTTGGGTAAAGAACTGCATACTCTTCTCTCTCTCTCTCTCT 1658
481 LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
1659 TTGCATCACAACCCAGATGTTAAATTTGGCTTGTATCCAGCATTCCAGGAAAGCGTAC 1718
501 TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPTrpAspTyrGly 520

```

DB 1719 TATTGGTTAAGAACTCTGATGAGCAGCAGTAAAGTTGCGCGTGGTGGATTACGGT 1778
QY 521 TTTGlnIleGlyClyMetAlaAspArgThrLeuValAspAsnThrTrpAsnAsn 540
DB 1779 TACCAAAATGGTGGCATGGCAGACAGAACCACTTTAGTCGATATAACAACACCGTCGAACAAT 1838
QY 541 ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTrpGluIle 560
DB 1839 ACTCACATCGCATCGTTGGTAAGCCATGGCTCCCTGGAAGAAATCTTACGAAT 1898
QY 561 LeuLysGluHisAspValAspTrpValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
DB 1899 CTAAAGAGCATGATGTCGATTATGCTTGTGTCATCTTTGGTGTCTAAATGGTTGGT 1958
QY 581 GlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGlu 600
DB 1959 GGTGATGACATCAACAATCTTGTGGATGATCAGAAATAGCGAGGAATCTGGCCAGAA 2018
QY 601 GluIleLysGluArgTrpPheTrpThrAlaGluGlyLysTrpArgValAspAlaArgAla 620
DB 2019 GAGATAAAGAGCGTTATTTCTATACCGCAGAGGAGAAATACAGAGTAGATGCAAGGCT 2078
QY 621 SerGluThrMetArgAsnSerLeuLeuTrpIleMetSerTrpLysAspPheProGlnLeu 640
DB 2079 TCTGAGACCATGAGAACTGCTACTTTTACAGATGTCCTACAAAGATTTCCACAAATTA 2138
QY 641 PheAsnGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
DB 2139 TTCAATGGTGGCCAAAGCACTGACAGAGTGGTCAACAATGATCACACCATTAGACGTC 2198
QY 661 ProProLeuAspTrpPheAspGluValPheThrSerGluAsnTrpMetValArgIleTrp 680
DB 2199 CCACCATAGACTCTTCGACGAGTCTTTTACTTCGAAAACCTGGATGGTTAGAAATAT 2258
QY 681 GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
DB 2259 CAATTGAAGAGGATGATGCCAAGGTAGAACTTTGAGGAGCGTTGGTGAATTAACGAG 2318
QY 701 SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal 718
DB 2319 TCTTCTACGAAACCCAGAGGTCCATAAAGAGACCTGAAATAGGCTTGAGATC 2372
RESULT 2
ACC61293
ID ACC61293 standard; DNA; 2000 BP.
XX ACC61293;
AC ACC61293;
XX
DT 20-JUN-2003 (first entry)
XX
DE Gene sequence #SEQ ID 1368.
XX
KW Multimeric protein complex; eukaryote; drug target; diagnosis; gene; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN EP1258494-A1.
XX
PD 20-NOV-2002.
XX
PF 20-DEC-2001; 2001EP-00130253.
XX
PR 15-MAY-2001; 2001EP-00111774.
XX
PA (CELL-) CELLZONE AG.
XX
PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzloch M, Schultz JD, Superti-Furga GD;
XX
DR WPI; 2003-250078/25.
DR P-PSDB; ABR53251.
XX
PT New isolated protein complexes useful for diagnosing a disease or

disorder, or as a target for an active agent of a pharmaceutical.
PT Preferably a drug target in the treatment or prevention of disease or
PT disorder.
XX
XX
XX Disclosure; SEQ ID NO 1368; 17pp + Sequence Listing; English.
XX
CC The invention relates to multiprotein complexes from eukaryotes. Proteins
CC of the invention and DNA sequences encoding them are given in records
CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC obtainable by using a protein as a bait and isolating the set of proteins
CC which is attached thereto from cells. Such protein complexes may comprise
CC up to 30 distinct proteins. Protein complexes of the invention are useful
CC for diagnosing a disease or disorder, or as a target for an active agent
CC of a pharmaceutical, preferably a drug target in the treatment or
CC prevention of a disease or disorder. Note: The sequence data for this
CC patent is not represented in the printed specification, but is based on
CC sequence information supplied by the European Patent Office. The complete
CC document is available on CD-ROM.

XX Sequence 2000 BP; 461 A; 459 C; 453 G; 627 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2000
Score: 3488.00 Matches: 665
Percent Similarity: 99.85% Conservative: 0
Best Local Similarity: 99.85% Mismatches: 1
Query Match: 92.74% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-6 (1-718) x ACC61293 (1-2000)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
DB 1 ATGGGATTCGACCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 60
QY 21 PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
DB 61 TTCTGTTGGCGAATTTTGGGGCTGCCATATCATCATCGTTGGTGGCAGTCATCAATTTGAG 120
QY 41 SerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTrpLeuValAsn 60
DB 121 TCTATATCCATGAATTCGACCCCTGGTTCATATATAGGGCTACCAATATCTCGTCAAC 180
QY 61 AsnSerPheTrpLysPheLeuAsnTrpPheAspArgThrTrpTrpProLeuGlyArg 80
DB 181 AATTCGTTTACAAAGTTTGAACCTGGTTGACACCGTACTGGTACCCCTCGGAGG 240
QY 81 ValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
DB 241 GTTACTGGAGGAGCTTTATATCTCTGGTTTGTATGACGACTAGTGGCTTCATCTGGCAGCC 300
QY 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
DB 301 CTGGCAACTGGTTGGCTGGCCATTCGACATCAGAAACGTTTGTGTGCTATTTCGCCCA 360
QY 121 LeuPheSerGlyValThrAlaTrpAlaThrTrpGluPheThrLysGluIleLysAspAla 140
DB 361 CTATTTCTGGGGTCACCGGCTGGCGACTTACGAATTTACGAAAGAGATTAAAGATGCC 420
QY 141 SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTrpIleSerArgSer 160
DB 421 ACCGCTGGGCTTTTGGCTGCTGGTTTATAGCATTGTCCCCGGTTATATATCTAGATCA 480
QY 161 ValAlaGlySerTrpAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
DB 481 GTGGCGGGGCTCTACGATAATGAGGCCATTCGCCATTAATTAATTAATTAATTAATTAAT 540
QY 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuPhe 200
DB 541 TTTTGGATTAAAGCCCAAAAGACTGGCTCTATCATGCACGCAACGTTGTCAGCTTTATTC 600
QY 201 TyrPheTrpMetValSerAlaTrpGlyGlyValPheIleThrAsnLeuIleProLeu 220

CC for the treatment or prevention of a disease or disorder. This sequence
CC corresponds to a gene of the invention. (Note: the sequence data for this
CC patent did not form part of the printed specification but was obtained
CC from the EPO in electronic format).

Sequence 2000 BP; 461 A; 459 C; 453 G; 627 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------|---------|------------------------|--------|
| | 0 | Length: | 2000 |
| Pred. No.: | 685 | Matches: | 685 |
| Score: | 3488.00 | Percent Similarity: | 99.95% |
| | | Best Local Similarity: | 99.85% |
| | | Mismatches: | 1 |
| | | Indels: | 0 |
| | | Gaps: | 0 |
| | | DB: | 12.74% |
| | | | 90.0 |

US-10-028-384-6 (1-718) x ADK63659 (1-2000)

[illegible]

1921 TTCAATGGTGGCCAGCCTAGTACAGAGTGGCTCAACAAATGATACACCATAGAGCTC 1980
 661 ProProteinasePyrPhe 666
 1981 CCACCATAGACTACTTC 1998

RESULT 4
 ABZ32036
 ID ABZ32036 standard; DNA; 2256 BP.
 XX AC ABZ32036;
 XX DT 30-JAN-2003 (first entry)
 XX DE Candida albicans essential gene SEQ ID NO 6323.
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX OS Candida albicans.
 XX WO200253728-A2.
 XX PD 11-JUL-2002.
 XX PF 26-DEC-2001; 2001WO-US049486.
 XX PR 29-DEC-2000; 2000US-0259128P.
 XX PR 20-FEB-2001; 2001US-00792024.
 XX PR 22-AUG-2001; 2001US-0314050P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 XX WPI; 2002-566694/60.
 XX DR P-PSDB; ABP73486.
 XX PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX PS Claim 37; SEQ ID NO 6323; 167pp + Sequence Listing; English.
 XX CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX SQ Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 116e-272 Length: 2256
 Score: 2738.00 Matches: 509
 Percent Similarity: 81.22% Conservatives: 88
 Best Local Similarity: 69.25% Mismatches: 110
 Query Match: 72.80% Indels: 28
 DB: 6 Gaps: 3

US-10-028-384-6 (1-718) x ABZ32036 (1-2256)

QY 9 LeuSerValPheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAla 28
 DB 52 GTTGAACGATTAGAGATTTTATTAAAGCTTATTATTATATCATGATGAGGTGCAGCT 111
 QY 29 IleSerSerArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspPro 48
 DB 112 ATTTCTTCTCGTTTATTTTCGGTATTCGATTGAAAGTATTATTATCATGATTCGATCCT 171
 QY 49 TrpPheAsnTyrArgAlaThrLysTyrLeuValIleAsnSerPheTyrLysPheLeuAsn 68
 DB 172 TGGTTCAATTTCCGAGCAACCAATATTATGCTACTCTTCCTTTTATGAATTTTGAAT 231
 QY 69 TrpPheAspArgThrTyrProLeuGlyArgValThrGlyGlyThrLeuTyrPro 88
 DB 232 TGGTTTGGATGATAGACTTGGTACCATTGGGAGAGTCACTGGTGGTACTTATATCCC 291
 QY 89 GlyLeuMetThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuPro 108
 DB 292 GGTTAATGGTGACTTCAGGTGCCATTTGGCATATTTTACGTGATGGTTTGGCTTACCC 351
 QY 109 IleAspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrp 128
 DB 352 GTTGATATTAGAAATATTGTGTTTATTAGCACCAAGTTTCTCGGATTAACGTCAATT 411
 QY 129 AlaThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaGly 148
 DB 412 TGTACTTATTTTGTGCTAAAGAAATGAAGGATTTCTAGTCAGGATTTATGGCAGCTATA 471
 QY 149 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 168
 DB 472 TTATGGGATTTGCCCGAGGTATATTTCAAGATCAGTGGCTGTTCTTATGATTAATGAA 531
 QY 169 AlaIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThr 189
 DB 532 GCRAATGGCAATTTACTTTTAAATGGAACATTTTATTTCTGGATTAATCAATGAAATG 591
 QY 189 GlySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrp 208
 DB 592 GGTTCAGTTTCTATGCCACATTTGACAGCATTTATTTCTATATGGTTAGTGTGG 651
 QY 209 GlyGlyTyrValPheIleThrAsnLeuLeuProLeuHisValPheLeuLeuLeuMet 228
 DB 652 GGTGGATATGTTTTCATTAACCAATTTGATTCATACACGTAATTTGCTTGAATTTTCATG 711
 QY 229 GlyArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrVal 248
 DB 712 GGTGCTTAAATGCCAATTTTACACTGCTTATACATGATGATGCTTGGTACTTGT 771
 QY 249 AlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAla 268
 DB 772 GCATCAATGCAGATTCCTCCATTCGTTGGGTTTTTACCATTAGATCAATGATCATCGCT 831
 QY 269 AlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGln 288
 DB 832 GCATTAGAGATTTTGGATTGTACAAATAGTGGCTTTGGTGGATTATGTTAATCAAAA 891
 QY 289 IleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeu 308
 DB 892 GTTCCAAACCAAAATTAATATTCCTGTAGTATTCCTATGTTGTGTGGATTA 951
 QY 309 GlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuLeuAlaProTyrThrGlyArg 328

Db 952 GGTATTGGTGGATTATTTGGATTAACAGCAATGGGTTGGATTGCTCTCTGGACAGGTAGA 1011
 Qy 329 PheTyrSerLeuThrAspThrAsnTyrAlaLysHisIlePheProIleAlaSerVal 348
 Db 1012 TTTTATTCTTATGGGATACAAATATGACCAAGATTCATATCCAAATATTGCTCTGTT 1071
 Qy 349 SerGluHisLeuProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrp 368
 Db 1072 TCGAACAATCAACCTACTGCTGGCAGCAATCTTTTCGATCTAGTAGCTTATTGG 1131
 Qy 369 LeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheVal 388
 Db 1132 TTATTCGCCGCTGCTATCTATTATGTTTCCAGAAATTCGAAGGATGAACAGCTTTTCA 1191
 Qy 389 IleAlaTyrSerValLeuLysSerTrpPheAlaGlyValMetValArgLeuMetLeuThr 408
 Db 1192 ATCATTTACAGTATGTTGTTCTTATTGCTGGTGTCATGGTAAGATTGATGTTGACT 1251
 Qy 409 LeuThrProValIleCysValSerAlaValAlaLeuSerLysIlePheAspIleTyr 428
 Db 1252 TTGACTCCAGTCATTTGTTGCTGCAGCAATTCCTTATCTAAATGTTTGTATGCTAT 1311
 Qy 429 LeuAspPhe----- 431
 Db 1312 TTGGACATGTTGATTGTTCTACTCGAAGGTTGGAAAGTATGATGACGTAGTGAC 1371
 Qy 432 -----LysThrSerAspArgLysTyrAlaLysProAlaAlaLeu 445
 Db 1372 GAATCCAAAGAGCTCAACCAAAATCAAGTTCACAGTTTCCAAAT-----GCTGGATAT 1425
 Qy 446 LeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyrLeuPheValPheHis 465
 Db 1426 TTCTCAAAAGTTTGGTTTACTGACATTTACATTTTACCTTTCTACTTTGTTTACAT 1485
 Qy 466 SerThrTrpValThrArgThrAlaTyrSerProSerValValLeuProSerGlnThr 485
 Db 1486 TGACTTGGGTAAACATCGAATGCTTATTCATCACCATCAGTTGTTTAGCATCCAGAAAC 1545
 Qy 486 ProAspGlyLysLeuAlaLeuAspPheArgGluAlaTyrTyrTrpLeuArgMet 505
 Db 1546 CCAGATGGCTCACACATATCATTTGATGATTATAGAGAAGCTATTACTGTTAAGATG 1605
 Qy 506 AsnSerAspGluAspSerLysValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGly 525
 Db 1606 AATACACCAAGATCCAAAGTTATGGCTGCTGGGATTATGGTTATCAATCGGGGT 1665
 Qy 526 MetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIle 545
 Db 1666 ATGGCTGATAGAACACACACTTGTGTGATAACAATACATGGAATAACACACATATGCCACT 1725
 Qy 546 ValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAsp 565
 Db 1726 GTTGGTAAGCAATGTTCCCTGAAGATGTGCTATGAATTTTGAGACAACACGAT 1785
 Qy 566 ValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyAspAspIleAsn 585
 Db 1786 GTTGATTATGTTAGTTATTTGGAGGTTATTGGGTTATCTGCTGGTATGATATTAAAC 1845
 Qy 586 LysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArg 605
 Db 1846 AATTTCTTATGGATGTAGAAATGCTGAAGGATCTGCTGCTGATGAATCAAGAAAGA 1905
 Qy 606 TyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAlaSerGluThrMetArg 625
 Db 1906 GACTACTTTACTGCCGAGGAGAATAAAGTGGATAAAGATGCATCATCTGCCAATGAAG 1965
 Qy 626 AsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGln 645
 Db 1966 AATCTTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2025
 Qy 646 AlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyr 665
 Db 2026 GGTGTTGATAGATTAGAAACCAACCAATCCCGAGCAATGAAGTACCCGAAATTTGAATGTT 2085

Qy 666 PheAspGluValPheThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAsp 685
 Db 2086 GTTGAAGAAGCCTTTCACATCAGAAATTTGGATTGTGAGATTTACAAAGTTAAAGATTG 2145
 Qy 686 AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer----- 702
 Db 2146 GATTAATGTTGGTAGAGATTTACATCAAGCTACTGCTTTTGAAGATTCATCATCCGCACT 2205
 Qy 703 ThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArg 717
 Db 2206 TCCAAAAGAAACAGATCCATAAGAGACCTAAATTTGGAAGTAGA 2250

RESULT 5

ABT20789
 ID ABT20789 standard; DNA; 2232 BP.

XX AC ABT20789;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene #3147.

XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection; cancer; contamination; biofilm; antibody; immune response; ds.

XX OS Aspergillus fumigatus.

XX PN WO200286090-A2.

XX PD 31-OCT-2002.

XX PP 23-APR-2002; 2002WO-US013142.

XX PR 23-APR-2001; 2001US-0285697P.

XX PR 27-APR-2001; 2001US-0287086P.

XX PR 05-JUN-2001; 2001US-0295890P.

XX PR 09-JUL-2001; 2001US-0303899P.

XX PR 31-AUG-2001; 2001US-0316362P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX DR WPI; 2003-093124/08.

XX PT New purified or isolated nucleic acids of essential genes of Aspergillus fumigatus, useful for treating or preventing infections by A. fumigatus, or for treating a non-infectious disease in a subject e.g. cancer.

XX PS Disclosure, Page: 175pp; English.

XX CC The invention relates to novel purified or isolated nucleic acids of essential genes of Aspergillus fumigatus. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as A. fumigatus, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by A. fumigatus, or to prevent or inhibit formation on a surface of a biofilm comprising A. fumigatus. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of A. fumigatus to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparisons with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively

determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organism invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This polynucleotide sequence represents one of the essential genes of *Aspergillus fumigatus* of the invention

Sequence 2232 BP: 473 A: 605 C: 545 G: 609 T: 0 U: 0 Other:

| | | | |
|------------------------|-----------|---------------|------|
| Alignment Scores: | 1.23e-240 | Length: | 2232 |
| Pred. No.: | 2428.50 | Matches: | 460 |
| Score: | 77.68% | Conservative: | 97 |
| Percent Similarity: | 64.1% | Mismatches: | 137 |
| Best Local Similarity: | 64.57% | Indels: | 23 |
| Query Match: | 8 | Gaps: | 6 |
| DB: | | | |

US-10-028-384-6 (1-718) X ABT20789 (1-2232)

| | | | |
|----|-----|---|-----|
| Qy | 15 | IleLeuLysLeuValIlePheValAlaIlePheGlyAlaIleSerArgLeuPhe | 34 |
| Db | 61 | CTTCGGGATTATTATCTCTGTGTACATTTCAGCAGCAGCAGTGGCTGACATCTTC | 120 |
| Qy | 35 | AlaValIleLysPheGluSerIleIleHisGlyPheAspProTrpPheAsnTrpArgAla | 54 |
| Db | 121 | ACGGTTATCGGCTTCGAGAGTATCATCCAGAGTTGACCCCGTGGTTCACATTCGAGCA | 180 |
| Qy | 55 | ThrLysTrpLeuValAsnSerPheTrpLysPheLeuAsnTrpPheAspArgThr | 74 |
| Db | 181 | ACAAAATACTTAGTACAGAATGGTTCTATAGCTTTTGGGATTGGTTGATGACGGAAC | 240 |
| Qy | 75 | TrpTrpProLeuGlyArgValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSer | 94 |
| Db | 241 | TGGCATCTCTGGGACGGTGCACCGGTGGCAGCTTATATCCCGTCTCATGGTGACGAC | 300 |
| Qy | 95 | AlaPheIleTrpHisAlaLeuArgAsnTrpIleuGlyLeuProIleAspIleArgAsnVal | 114 |
| Db | 301 | GGCGTGATCTACCATATCTCGCA--TTCCTTACTATCCCGCTCGATATTCGCAACATC | 357 |
| Qy | 115 | CysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTrpGluPheThr | 134 |
| Db | 358 | TGGGTCTACTGGCGCCAGGATTCTCGGCGCTGACTGCATTGGCAATGTACTTGTGTACA | 417 |
| Qy | 135 | LysGluIleLysAspAla---SerAlaGlyLeuLeuAlaIleGlyPheIleAlaIleVal | 153 |
| Db | 418 | TCGAGATGTCCTCTCCCATCTCGCAGGTCTTCTTCGACAGCTTTCATGGGAATCGCC | 477 |
| Qy | 154 | ProGlyTrpIleSerArgSerValAlaGlySerTrpAspAsnGluAlaIleAlaIleThr | 173 |
| Db | 478 | CTGGTTACATCTCCCGATCAGTTGCTGGAAGCTACGATTAACGAGCGATTGCCATCTTT | 537 |
| Qy | 174 | LeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHis | 193 |
| Db | 538 | CTGCTGTGTTGCACATTTCTTTCTATGGATCAGGCTGTCAAAATGGGTCTATCATGTGG | 597 |
| Qy | 194 | AlaThrCysAlaAlaLeuPheTrpPheTrpMetValSerAlaTrpGlyGlyTrpValPhe | 213 |
| Db | 598 | GGAGCGCTGACCGCACTATTCTACGGCTACATGGTGTGGCATGGGTGGGTATGTCCTTC | 657 |
| Qy | 214 | IleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTrpSerSer | 233 |
| Db | 658 | ATTACGAACCTGATCCCGCTGCAGTCTTTGTCCCTCTGTGCAATGGGTAGATACAGCACT | 717 |
| Qy | 234 | LysLeuTrpSerAlaValThrThrTrpTrpAlaIleGlyThrValAlaSerMetGlnIle | 253 |
| Db | 718 | CGCATCTACATTAGCTATACCACTGATGGCTGGGACTTTGGCTAGCATGCAGATT | 777 |
| Qy | 254 | ProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaIleuGlyValPhe | 273 |
| Db | 778 | CCCTTCGTGGGATTTTGCCTATCCGAACAGCAGCACCATGTCCCGCTTGGGTGTCTTC | 837 |
| Qy | 274 | GlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLys | 293 |
| Db | 838 | GGCTGTCTTCACTGTGGCTTCTCGCAGATTTCTCGCAGCTTGTCTTCACAGCAAGCA | 897 |

| | | | |
|------|----|--|------|
| 294 | QY | PhelysValilleMetMetValSerLeuPheLeuLeuValLeuValLeuGlyValValGlyLeu | 313 |
| 898 | DB | TTCCAGACACTTTCGACCGGCATGATCTTCATCAACCTTCGGTCTCGGTTTGGTTGGGCTA | 957 |
| 314 | QY | SerAlaLeuThrTyrMetGlyLeuLeuAlaProTTPThrGlyArgPheTyrSerLeuTyr | 333 |
| 958 | DB | GTGTGTTCTGACTGTCGACGGGAGTGATCGCTCTCTGGAGCGGCCGATCTACTCTCTGTGG | 1017 |
| 334 | QY | AspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnPro | 353 |
| 1018 | DB | GACACTGCTATGCCAAATCCCAATCCCATTCATTCGCTCAGTCTCGGAACACCCAGCCC | 1077 |
| 354 | QY | ValSerTTPProAlaPhePhePheAspThrHisPheLeuLeuTTPLeuPheProAlaGly | 373 |
| 1078 | DB | ACCGTGTGGCCAGGTTCTTCTTCGATCTGAACCTCTCTGATCTGGCTTTTCCCGCAGGT | 1137 |
| 374 | QY | ValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerVal | 393 |
| 1138 | DB | GTCTACATGTGCTTCCGTGACCTCAAGGACGACATGCTTCGTCTATTACTACTCGGTC | 1197 |
| 394 | QY | LeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIle | 413 |
| 1198 | DB | CTTCGAGCTACTTCGCCGGTGTTATGTGTCCAGCTAAATGCTGACCTTGACCCCTATTGTG | 1257 |
| 414 | QY | CysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu | 429 |
| 1258 | DB | TGTGTTCGGCTGCTCTGGCGCTGTCTCCATCTTCGACACCTATATGGCGACTACCCCTC | 1317 |
| 430 | QY | -----AspPheLysThrSerAsp----- | 435 |
| 1318 | DB | CGGACACAGCGCTCGAAGGAAACGAATGAAGACTCGTCTTCACAACTCTTCGCTCA | 1377 |
| 436 | QY | ---ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySer | 454 |
| 1378 | DB | GTTCCGAAGCCCAATGTTGGAATTCACCTCCCATGTTTCTAAGATATTATGACGGCGTCT | 1437 |
| 455 | QY | PheIlePheTyrLeuTyrLeuPheValPheHisSerThrTTPValThrArgThrAlaTyr | 474 |
| 1438 | DB | GTTGTGCTGTACCTGTCTCTGTTTGTGGCACTGCACCTCGGTTACATCGAATGCATAC | 1497 |
| 475 | QY | SerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAsp | 494 |
| 1498 | DB | TCTTCTCTCCGTGTCCTGGCTAGTCGGATCGCTGACGAAGCCCAATACATCATGTGAC | 1557 |
| 495 | QY | AspPheArgGluAlaTyrTyrTTPLeuArgMetAsnSerAspGluAspSerLysValAla | 514 |
| 1558 | DB | GATTATCGTAGGCTTACTACTGGCTTCGTCTCAGAATATCTCTCAGAACGCCAAAATCATG | 1617 |
| 515 | QY | AlaTTPTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAsp | 534 |
| 1618 | DB | TCATGTTGGGATTATGGGTATCAATCGTGGCATGGCGGACGCCCAACTTGGTTGAC | 1677 |
| 535 | QY | AsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu | 554 |
| 1678 | DB | AACAAACCTCGAACACACCCATATGTCGCTGTGTAGGCGATGAGCTCACCGGAG | 1737 |
| 555 | QY | GluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGly | 574 |
| 1738 | DB | GAAGTCAGTACCCATCTCTCCGCCAGCATGATGCGATTACGTGCTGGTGTTCGGT | 1797 |
| 575 | QY | GlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTTPMetIleArgIleSer | 594 |
| 1798 | DB | GGTCTGTAGGTATTCTCGCGATGCAATTACAATAATCTTATCGATGGTCCGTATGCC | 1857 |
| 595 | QY | GluGlyIleTTPProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGlyTyr | 614 |
| 1858 | DB | GAAGGTATCTGGCCCATGAGTATAAGAGCGGGGACTTCTTTACTGCACCGCGTGAATAT | 1917 |
| 615 | QY | ArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyr | 634 |
| 1918 | DB | CGTGTCAACATGAGCGACCCCACTATGCGCAACAGCTTGTGGTATAAAATGCTTAT | 1977 |

635 LysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMet 654
 1978 TACAATTTCAACTCTCTCTCCGTCGGCCCAAGCTGTCACCGGTCGTTGGGTCAAAA 2037
 655 IleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAsn 674
 2038 CTT---CCACAGAAGGCCCTCAGCTCTCTACCTGCAAGAAGCTTTACAGAGCAGAAC 2094
 675 TrpMetValArgIleTyrGlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAsp 694
 2095 TGGATCATTCGTATCTACAGGTCAGGATCTTGACAACTTGGCCGAGCACCACACAG 2154
 695 ValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLysArg 711
 2155 GCTGTTGCTTCGACAAA---GGTCTCAAGAAAAGCCGAGTACAAAAGAGG 2202

RESULT 6

ADB69900
 ID ADB69900 standard; DNA; 2466 BP.

AC ADB69900;

XX 04-DEC-2003 (first entry)

XX C. neoformans open reading frame SEQ ID NO:2305.

DE ds; gene; fungicide; gene therapy; infection.

XX Cryptococcus neoformans.

XX WO2003052076-A2.

XX 26-JUN-2003.

XX 17-DEC-2002; 2002WO-US040225.

XX 17-DEC-2001; 2001US-0341261P.

XX (ELIT-) ELITRA PHARM INC.

XX Zamudio C, Eroshkin AM;

XX WPI; 2003-533017/50.

XX P-PSDB; ADB70261.

XX New nucleic acid, useful for preparing a composition for treating an

XX infection caused by Cryptococcus neoformans.

XX Claim 2; SEQ ID NO 2305; 136pp; English.

XX The invention relates to a novel purified or isolated Cryptococcus
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 1.99e-227 | Length: | 2466 |
| Score: | 2301.50 | Matches: | 435 |
| Percent Similarity: | 73.05% | Conservative: | 107 |
| Best Local Similarity: | 58.63% | Mismatches: | 155 |
| Query Match: | 61.13% | Indels: | 45 |
| DB: | 10 | Gaps: | 6 |

US-10-028-384-6 (1-718) x ADB69900 (1-2466)

QY 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArg 32
 DB 223 GAACGCTTTGGCTTCTCATCTCTCGGTAAATATGCGGTCTGCGATTTGAAGTCGA 282
 QY 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsnTyr 52
 DB 283 TTGTTTGGCGTCATCAGATTGCAATCTGTCTATCCACGAATTTGACCCCTGGTCAACTAC 342
 QY 53 ArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAsp 72
 DB 343 CGAGCTCGAAAGCTCTTGTTTACAGGGTTTCTACGAGTTCTTGGAACTGCTTTGACCCC 402
 QY 73 ArgThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThr 92
 DB 403 TCCGCTTGGTACCTCTCGGAGAACTGCTGGTACCACTGCTATCTGCTGGCTTGGTGC 462
 QY 93 ThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArg 112
 DB 463 ACGTCTGGACTGATTTGGCATGCTCTTCCGGCA---ATCAATATGCCCGGACATTCGC 519
 QY 113 AsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGlu 132
 DB 520 AATGCTCTGCTCCTTGCACCTGGATTTTCTGGATTGACTGCTGGCGGACTTATCTT 579
 QY 133 PheThrLysGluIleLysAspAlaSerAlaGlyLeuAlaAlaGlyPheIleAlaIle 152
 DB 580 TTCACCACTGAATGCTTACACCATGAGTGTCTATTGGCGCGCTTTCATGGCATT 639
 QY 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIle 172
 DB 640 GTACCGGATACATCTCTCGATCTGTCCGCGGTTCTTATGACAAACGACCATTTGCCATC 699
 QY 173 ThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMet 192
 DB 700 TTCTCTTGATGAGCTCTCTACTCTTGGATTAGCCCGTCAAAACCGTAGCTCATTT 759
 QY 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVal 212
 DB 760 TGGGATATGATCACTGCTCTGTTCTACGGGTGGATGGTTCTGCTCATGGGGTGGTTACGTT 819
 QY 213 PheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSer 232
 DB 820 TTCAATCACCACATGATTCATTCGACGCTTGTCTCATTTGTCATGTCATGGCAGGTCAC 879
 QY 233 SerLysLeuTyrSerAlaTyrThrTyrTyrAlaIleGlyThrValAlaSerMetGln 252
 DB 880 AACCGGCTTTATACCGCTTACTCTCTCTGTTATGCTATTGGAACCTATCGCTCCATCAG 939
 QY 253 IleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyVal 272
 DB 940 GTCCCTTTGGAGTTCTCTCCCATCCGAACCTCTGAGCACATGCGGCGCTTGGTGT 999
 QY 273 PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla 292
 DB 1000 TTCGGCTTGPACAGCTGATCGGATTCGTGAAAGTCTGTCGACGACTCGTCTGCTGCAAG 1059
 QY 293 LysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValGly 312
 DB 1060 CAATTCAGCTCTCTCAAGCTTTGCTGCGCGGTATTCGCTCAGTTTGTGCTGC 1119
 QY 313 LeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeu 332
 DB 1120 CTCGTCATTTGACTTTCTCTGGATGGATCGCCCTCTGCTGGAAGATTTTATCTCTT 1179
 QY 333 TrpAspThrAsnTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGln 352
 DB 1180 TGGGATAGTGGCTATGGAAGGTCACATGCCATTTATTCCTCTCCGACACACAG 1239
 QY 353 ProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpPheProAla 372
 DB 1240 CCCACCGCTTGGCGCTCATTTCTACTTGTACCTCGAAATGCTTATCTCTTTTCCCTGCC 1299
 QY 373 GlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSer 392

Alignment Scores:

Pred. No.: 1,458-223 Length: 2603
 Score: 2264.50 Matches: 461
 Percent Similarity: 66.31% Conservative: 98
 Best Local Similarity: 54.69% Mismatches: 135
 Query Match: 60.21% Indels: 150
 DB: 8 Gaps: 11

US-10-028-384-6 (1-718) x AB20191 (1-2603)

| | | | | | | | |
|----|-----|--|-----|----|------|---|------|
| Qy | 15 | IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe | 34 | Db | 957 | TAGATACAGCACTCGCATCTACATTAGCTATACCAATGGTATGCGTGGGAGCTTTGGC | 1016 |
| Db | 61 | CTTCGCGGATATTATCTCTGTTACCATTCAGCAGCAGCAGTTCGACGCTGACTTTC | 120 | Qy | 249 | aserMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisValAlaAl | 269 |
| Qy | 35 | AlaValIleLys----- | 38 | Db | 1017 | TAGCATGCAGATCCCTTCGTCGGATTTTGGCTATCCGAAACAGCGACACATGTCGC | 1076 |
| Db | 121 | AGCGTTATCCG-TAAGTCATCAGGGAATCAAGTTCGTAATTTGTTAATCATCATCA | 179 | Qy | 269 | aleu----- | 271 |
| Qy | 39 | -----PheGluSerIleIleHisGlu----- | 45 | Db | 1077 | CTTGGTAGGTCCCTTGGCAAGTTCAAATGGCCGTTTCAGAGCTGACATTTTTCGAG | 1136 |
| Db | 180 | AATTGACTGATCATTCTTCTACAGGCTTCGAGAGTATCATCCAGAAATGTAAGTAT | 239 | Qy | 271 | lyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerT | 291 |
| Qy | 45 | ----- | 45 | Db | 1137 | GTFCTTCGGCTGCTTCAGCTTGTGGCTTCGCGAGTTTTCGCGAGCTTGTCCCA | 1196 |
| Db | 240 | AGTCAATTGTATACCTACGGCTTCGGCTTTAACGAGCCTCGCTTGGACGAATTGATA | 299 | Qy | 291 | hrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyVal | 311 |
| Qy | 60 | AsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAsp----- | 72 | Db | 1197 | GCAAGCAGTTCAGAGACTTTCGACCCCATGATCTTCATCATCTTCGGTCTCGGTTGC | 1256 |
| Db | 360 | CAGAAATGTTCTATAGCTTTTGGGATTGGTTTGATGACCGTATGCTTCCCGACGCA | 419 | Qy | 311 | alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrS | 331 |
| Qy | 72 | ----- | 72 | Db | 1257 | TTGGCTAGTTTGTCTGACTGTGACGGGAGTATCGCTCTTGGAGCGCGGATTCATCT | 1316 |
| Db | 420 | CAAGAGTGATGAAGCAACAGCTACCTGCTCGTGGAGTGTCTTAGGCTAATCGGGTCC | 479 | Qy | 331 | erLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluH | 351 |
| Qy | 73 | -----Arg-ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLe | 90 | Db | 1317 | CCTTGTGGGACACTGGCTATGTCACAAATCCACATTCATTCATCTCGTCTCGGAAC | 1376 |
| Db | 480 | CAACATAGGAACATCGCATCTCTGGAGCGTGTACCGGTGGCAGCTTATATCCGGTCT | 539 | Qy | 351 | isGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheP | 371 |
| Qy | 90 | uMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs | 110 | Db | 1377 | ACCAGCCACCGCTGGCCAGCGTTCCTTCGATCGAACTTCCTGATCTGCTTTTCC | 1436 |
| Db | 540 | CATGGTGACGGCGGCGATCATCTTCGCA--TTCCTTACTATCCCGCGCA | 596 | Qy | 371 | roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla | 391 |
| Qy | 110 | pileArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlath | 130 | Db | 1437 | CGGAGGTGTCTACATGTGCTTCGTCACCTCAAGAGCAGCATGCTTCGTCATTATCT | 1496 |
| Db | 597 | TATTGCAACATCTCGCTCTACTCGCGCCAGGATTCCTCGGCTGACTGCAATTGGCAAT | 656 | Qy | 391 | yrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThP | 411 |
| Qy | 130 | rTyrGluPheThrLysGluIleLysAla---SerAlaGlyLeuLeuAlaAlaGlyPh | 149 | Db | 1497 | ACTGGTCTTTCGAGAGTACTTTCGCGGTGTATGGTCCGACTTAATGCTGACCTTGACC | 1556 |
| Db | 657 | GTACTGTGACATCCGAGATGTCTTCCTTCGCCATCTGCAGGTCTTCTTCGACAGCTT | 716 | Qy | 411 | roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu---- | 429 |
| Qy | 149 | eileAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAl | 169 | Db | 1557 | CTATTGTGTGTGTCGCGTCTCTGCGCTGTCTGCTCATTCGACACCTATATGCGGA | 1616 |
| Db | 717 | CATGGGAATCGCCCTGGTTACATCTCCGATCATCTGCTGGAAGCTACGTAACGAAGC | 776 | Qy | 430 | -----AspPheLysThrSerAsp----- | 435 |
| Qy | 169 | alleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGl | 189 | Db | 1617 | CTACCTCCCGACACCGAGCTCTGAAGCGAAACGAATGAAGACTCGTCTTCACAACTTC | 1676 |
| Db | 777 | GATTGCAATCTTCTGCTGTGTTCACATCTTTTCTATGGATCAAGGCTGCAAAATGG | 836 | Qy | 436 | -----ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValS | 452 |
| Qy | 189 | ySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGl | 209 | Db | 1677 | TTGCTCATGTTCCGAAAGCCCAATGTTGGAATCACCTCCCATGTTCTTAAGATTATAGTCA | 1736 |
| Db | 837 | GTCTATCATGTGGGAGCGCTGACCGCACTATTCTACGGCTACATGTTGTCGCGATGGG | 896 | Qy | 452 | erGlySerPheIlePheTyrLeuPheValPheHisSerThrTrpValThrArgT | 472 |
| Qy | 209 | yGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuIleLeuMetGl | 229 | Db | 1737 | CGGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1796 |
| Db | 897 | TGGTATGCTTCATTAGAACCTGATCCCGCTGACGCTTTTCTCTCTCTGTCATGGG | 956 | Qy | 472 | hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla | 492 |
| Qy | 229 | yArgTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAl | 249 | Db | 1797 | ATGCATATCTTCTTCCTTCGTTGCTGTGCTAGTCGAGTCCCTGACGAAAGCAATACA | 1856 |
| | | | | Qy | 492 | euIleAspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerL | 512 |
| | | | | Db | 1857 | TCAITTAGCATATTCGTGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1916 |
| | | | | Qy | 512 | ysValAlaAlaTyrPheTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrL | 532 |
| | | | | Db | 1917 | AAATCATGCTCATGTGGGATTAAGGTATCAAAATCGGTGCGATGCGGACCGCCCACT | 1976 |
| | | | | Qy | 532 | euValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlas | 552 |
| | | | | Db | 1977 | TGGTITGACAAACACCTGGAAACACCATATTCGTCGCTGAGGTGGTAGGCGATGAGCT | 2036 |
| | | | | Qy | 552 | erProGluGlyLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValI | 572 |
| | | | | Db | 2037 | CACGCGAAGTACAGCTACCCCATCTCCCGCAGCATGATGTCGATACGTCGCTGCTGG | 2096 |

QY 572 lePheGlyGlyLeuIleGlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleA 592
 DB 2097 TGTTCGGTGGTCTGCTAGGTATTCCTGGCGATGACATTAACAAATCTTATGATGGTCC 2156
 QY 592 TGTTCGGTGGTCTGCTAGGTATTCCTGGCGATGACATTAACAAATCTTATGATGGTCC 612
 DB 2157 GTATCGCGAAGGTATCTGCGCGATGAGCTTAAGAGCGGACATCTTTACTGCACCG 2216
 QY 612 TGTTCGGTGGTCTGCTAGGTATTCCTGGCGATGACATTAACAAATCTTATGATGGTCC 629
 DB 2217 GTATCGCGAAGGTATCTGCGCGATGAGCTTAAGAGCGGACATCTTTACTGCACCG 2276
 QY 630 -----Tyr 630
 DB 2277 TCCCTTCCTCCCTCTTACTACTGCGTGCATCAGGTAAATCAAAATCTAGGTAT 2336
 QY 631 LysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
 DB 2337 AAAATGCTTATTACAAATTTCAACTCTCTCTCCGTCGGCCAAAGCTGCGACCGGTC 2396
 QY 651 ArgGlnMetIleThrProLeuAspValProLeuAspTyrPheAspGluValPhe 670
 DB 2397 CGTGGTCAAACTT---CCACAGAGGCGCTCTCTACTCTGAAAGAGCTTTC 2453
 QY 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnArg 690
 DB 2454 ACGAGCGAGAACTGATCTGATCTATCTCAAGGTCAAGGTCTTGACAACTTGGCGA 2513
 QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgSerIleLys 710
 DB 2514 GACCAACACAGGCTGTGCGCTTCGACAAA---GGTCAAGAAAAGCGGAGTCAAG 2570

711 Arg 711
 2571 AGG 2573

RESULT 8
 ABT17781
 ID ABT17781 standard; DNA; 3969 BP.

AC ABT17781;
 DT 16-APR-2003 (first entry)

DE Aspergillus fumigatus essential gene #139.

KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 cancer; contamination; biofilm; antibody; immune response; ds.

OS Aspergillus fumigatus.

PN WO200286090-A2.

PD 31-OCT-2002.

PF 23-APR-2002; 2002MO-US013142.

PR 23-APR-2001; 2001US-0285697P.

PR 27-APR-2001; 2001US-0287066P.

PR 05-JUN-2001; 2001US-0295890P.

PR 09-JUL-2001; 2001US-0303899P.

PR 31-AUG-2001; 2001US-0316322P.

XX (ELIT-) ELITRA PHARM INC.

XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of Aspergillus

PT fumigatus, useful for treating or preventing infections by A. fumigatus,

PT or for treating a non-infectious disease in a subject e.g. cancer.

XX PS The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterization, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention
 XX SQ Sequence 3969 BP; 964 A; 996 C; 953 G; 1056 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,7e-223 Length: 3969
 Score: 2264.50 Matches: 461
 Percent Similarity: 66.31% Conservative: 98
 Best Local Similarity: 54.69% Mismatches: 135
 Query Match: 60.21% Indels: 150
 DB: 8 Gaps: 11

US-10-028-384-6 (1-718) x ABT17781 (1-3969)

QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34

DB 698 CTTCTCGGATATTATCTCTGTGTACCATTCAGCAGCAGCAGTCCAGTCGACTCTTC 757

QY 35 AlaValIleLys----- 38

DB 758 AGCGTTATCCG-TAAGTCATCGAGGGAATCAAGTGTCTAAATTTGTTAATCTATCATCA 816

QY 39 -----PheGluSerIleIleHisGlu----- 45

DB 817 AATTGACTGATCATTCTTTCTTACAGGCTTCGAGAGTATCATCCAGCAATGTAAAGTAT 876

QY 45 ----- 45

DB 877 AGTCAATTTGTATACCTACGGCCTTCGGCCTTTAACGGACCTCGCTTGGACGAATTCATA 936

QY 46 -----PheAspProTrpPheAsnTyrArgAlaThrLysTyrLeuVal 59

DB 937 CTGATGCTCCGATATAGTTGACCCGTTGTTCACTTCGAGGCAACAAATACTTAGTA 996

QY 60 AsnAsnSerPheTyrLysPheLeuAsnTrpPheAsp----- 72

DB 997 CAGATGGTTTCTATAGCTTTTGGGATTTGGTGTATGATCACCCTATGTGTTCCCGACCGCA 1056

QY 72 ----- 72

DB 1057 CAAGAGTGCATGAGCAACAGGTACCTGTCTCTGGAGTGTCTTAGGCTAATCGGCGTCC 1116

QY 73 -----Arg-ThrTrpTyrProLeuGlyArgValThrGlyThrLeuTyrProGlyLe 90

DB 1117 CAACATAGAACATGCGATCTCTCTGGACGTGTCCCGTGGACGATTAATATCCCGTCT 1176

QY 90 uMetThrThrSerAlaPheIleThrHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs 110
 Db 1177 CATGGTACGAGCGCGGTGATACCATATCTTTGCGA---TTCCCTACTACTCCCGTCGA 1233
 QY 110 PileArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaTh 130
 Db 1234 TATTGCAACATCTCGGTCTACTGCGCCAGGATTCTCGGCTGACTGCATTGGCAAT 1293
 QY 130 rTyTCluPheThrIysGluIleIysAspAla---SerAlaGlyLeuAlaAlaGlyPh 149
 Db 1284 GTACTGTGTCATCCGAGATGCTCTCTCGCATCTGCGAGTCTTCTTGCAGAGCTTT 1353
 QY 149 eileAlaIleValProGlyTrpIleSerArgSerValAlaGlySerTyrAspAsnGluAl 169
 Db 1354 CATGGGAATCGCCCTGGTTACATCTCCGATCAGTTGCTGGAAGCTACGATAAGCAAGC 1413
 QY 169 aileAlaIleThrLeuLeuMetValThrPheMetPheTrpIleIysAlaGlyThrCl 189
 Db 1414 GATTGCCATCTTCTGCTTGTGTCACATCTTCTTATGGATCAAGGCTGCAAAATGG 1473
 QY 189 ySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGl 209
 Db 1474 GTCTATCATGTGGGAGCGGTGACCGCATCTATTCTACGGCTACATGGTGTGCGCATGGGG 1533
 QY 209 yGlyTyrValPheIleThrLeuIleProLeuHisValPheLeuLeuIleLeuMetGl 229
 Db 1534 TGGGTATGCTCTCATACGACCTGATCCCTTCGACGTTTGTCTTCTGTGCGATGGG 1593
 QY 229 yArgTyrSerSerIysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAl 249
 Db 1594 TAGATACAGCACTCGCATCTACATTAGCTATACCATGATGATGCGTGGGACTTTGGC 1653
 QY 249 aSerMetClnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
 Db 1654 TAGCATGCAGATCCCTTCGCGGATTTTGCCTATCCGAAACAGCGGCACCATGTCGCG 1713
 QY 269 aLeu-----G 271
 Db 1714 CTTGGGTAGGTCCTTTCGCAAGTTCAAATGGCGGCTTCAGGAGTCGACATTTTGCAG 1773
 QY 271 lyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValIysGlyGlnIleSerT 291
 Db 1774 GTGCTTCGGCGCTGCTCAGCTTGTGGCTTCGCGAGTTGTCCGAGCTTCGTTCCAA 1833
 QY 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValV 311
 Db 1834 GCAAGCAGTTCAGAGACTTCTACCGCCCATGATCTTCATCACCCTTCGGTCTCGGTTTCG 1893
 QY 311 alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrS 331
 Db 1894 TTGGCTAGTGTCTGACTGTGACGGGAGTGATCGCTCTTGAGCGGCGGATTTCTACT 1953
 QY 331 erLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleAlaSerValSerGluH 351
 Db 1954 CTTGTGGGACACTGGCTATGCCAAATCCACATTCCTCATCTGCTCAGTCTCGGAAC 2013
 QY 351 isGlnProValSerTrpProAlaPhePhePheThrHisPheLeuIleTrpLeuPheP 371
 Db 2014 ACCAGCCACCGCTTGGCGAGCTTCTTTCGATCTGAACTTCCTGATCTGCTTTTCC 2073
 QY 371 roAlaGlyValPheLeuLeuPheLeuAspLeuIysAspGluHisValPheValIleAla 391
 Db 2074 CGGCAGGTGTCTACATGTCTTCGCTGACCTCAAGGACGAGCATGTCTTCGTCATTATCT 2133
 QY 391 yrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuTrp 411
 Db 2134 ACTCGGCTCTCGGAGCTACTTCGCGGTGTATGGTCCGACTAAATGCTGACCTTGACCC 2193
 QY 411 roValIleCysValSerAlaAlaValAlaLeuSerIysIlePheAspIleTyrLeu--- 429
 Db 2194 CTATTGTGTGTGGCTGCTCTGCGCTGTGCTCCATTCTCGACACCTATATGGCGA 2253

RESULT 9

ABT19595

ID ABT19595 standard; DNA; 4603 BP.

XX

AC ABT19595;

QY 430 -----AspPheIysThrSerAsp----- 435
 Db 2254 CTACCTCCCGACACACGAGGTCTGAAGCGAARACGATGAGACTCGTCTTCACAACTC 2313
 QY 436 -----ArgLysTyrAlaIleLysProAlaLeuLeuAlaLysLeuIleValS 452
 Db 2314 TTCGCTCAGTTCGGAAGCCCAATGTGGAAATCACCTCCCATGTTTCTAAGATTATAGTGA 2373
 QY 452 erGlySerPheIlePheTyrLeuPheValPheHisSerThrTrpValThrArgT 472
 Db 2374 CGGGCTGTGTGCTCTACCTCTCTCTTGTTCGGCCTGACCTGGGTATACGTA 2433
 QY 472 hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 492
 Db 2434 ATGCATACTCTTCTCTCTCGTTGTCTTGGCTAGTGGATGCTGACGGAAGCCCAATACA 2493
 QY 492 euIleAspAspPheArgGluAlaTyrTrpLeuArgMetAsnSerAspGluAspSerL 512
 Db 2494 TCATTGACGATTAATCTGAGGCTTACTACTGCTTCTGCTGAGATCTACTCTCAGAACGCCA 2553
 QY 512 ysValAlaAlaIleTrpTyrGlyTyrGlnIleGlyMetAlaAspArgThrL 532
 Db 2554 AAATCATGTCTGCTGGATATGGTATCAAAATCGTGGCATGGCGGACCCCAACT 2613
 QY 532 euValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAla 552
 Db 2614 TGGTTGACAAACAACCTGGAACACACCCATATGCTACGTTGTAAGGCGATGAGCT 2673
 QY 552 erProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValI 572
 Db 2674 CACGCGAGGAAGTACGATACCCCATCTCCGCGCAGCATGATGCTGATTCAGTGTGGTGG 2733
 QY 572 lePheGlyGlyLeuIleGlyPheGlyGlyAspIleAsnLysPheLeuTrpMetIleA 592
 Db 2734 TGTTCGGTGTCTGCTAGGTATCTGGCGATGACATTAACAAATTCITATCGATGGTCC 2793
 QY 592 rgIleSerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGlu 612
 Db 2794 GTATCGCGAAGGTATCTGCGCGATGAGGTAAAGAGCGGACTTCTTTACTGCACGCG 2853
 QY 612 lyGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeu----- 629
 Db 2854 GTGATATCTGTGCGAGATGAGGACGCCCACTATGGCGAACAGCTTGATGATGAT 2913
 QY 630 -----Tyr 630
 Db 2914 TCCCTCTCTCCCTCTTATCTACTGCGCTGCATCACGCTAAATTCACAAATCTAGGTAT 2973
 QY 631 LysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
 Db 2974 AAATGCTTATTAACATTTCACTCTCTTCCGTCGGGCAAGCTGTCACCGCGCTC 3033
 QY 651 ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAspGluValPhe 670
 Db 3034 CGTGGGTCAAAACTT---CCACAGAAAGGCCCTCAGCTCTCTACACCTCGAAGAGCTTTC 3090
 QY 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArg 690
 Db 3091 ACAGCGAGAACCTGGATCATCTGTAICTACAGGTCAAGATCTTCAGAACCTTGGCCGA 3150
 QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgSerIleLys 710
 Db 3151 GACCACACCCAGGCTGTGCTTCGCAAAA---GGTCTCAAGAAAAGCGGAGTACAAG 3207
 QY 711 Arg 711
 Db 3208 AGG 3210

Db 2137 GTGTCCTCGCCTGCTAGCTTGTGGCTTTCGCGAGCTTGTGCTTCCAA 2196
QY hrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuValV 311
Db 2197 GCAAGCAGTTCAGAGACTTCTGACCGCATGCTTCATCACCTTCGGTCTGGTTGC 2256
QY 311 aGlyLeuSerAlaLeuThrTyMetGlyLeuIleAlaProThrGlyArgPheTyS 331
Db 2257 TTGGGTAGTTGTTGACTGTGACGGAGGATCGCTCTTGGAGCGCGGATCTTACT 2316
QY 331 erLeuTrpAspThrAsnTyAlaLysIleHisIleProIleIleAlaSerValSerGluH 351
Db 2317 CTTTGGGACACTGCTATGCAAAATCCACATTCCTCATTCCTGCTGCTCGGAAC 2376
QY 351 iGlnProValSerTrpProAlaPhePheAspThrHisPheIleIleTrpLeuPheP 371
Db 2377 ACCAGCCCACTTGGCGAGTCTTCTTCGATCTGAACTTCTGATCTGGCTTTTC 2436
QY 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla 391
Db 2437 CGGAGGTCTACATGTCTTCCTGACCTCAAGGACGAGCATGTCTTGTGCTATCT 2496
QY 391 yrSerValLeuCysSerTyPheAlaGlyValMetValArgLeuMetLeuThrP 411
Db 2497 ACTCGTCTTGGAGCTACTTGGCGGTGTATGCTCGGACTAATGCTGACCTTGACC 2556
QY 411 roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyLeu 429
Db 2557 CTATTGTGTGTGCGGTCTCTGCGCTGTGCTCCATCTCGACACCTATATGCGCA 2616
QY 430 -----AspPheLysThrSerAsp----- 435
Db 2617 CTACCTCCGACACAGGCTCGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 2676
QY 436 -----ArgLysTyAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleVal 452
Db 2677 TTCGCTCAGTTCGGAAGCCCAATGTGGAATCACCTCCATGTTTCTAAGATTATAGTGA 2736
QY 452 erGlySerPheIlePheTyLeuTyLeuPheValPheHisSerThrTrpValThrArgT 472
Db 2737 CGCGCTCTTGTCTACCTGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2796
QY 472 hrAlaTySerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 492
Db 2797 ATGCATACCTCTCTCTTCGTTGCTTGGCTAGTGGATGCTGACGGAAGCAATACA 2856
QY 492 euIleAspPheArgGluAlaTyTyTrpToLeuAtcMetAsnSerAspGluAspSerL 512
Db 2857 TCATTGACGATTATCGTAGGCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2916
QY 512 ysValAlaAlaTrpTrpAspTyGlyTyGlnIleGlyGlyMetAlaAspArgThrL 532
Db 2917 AAATCATGTCATGTTGGATTATGCTATCAATCGGTGCTGCTGCTGCTGCTGCTGCT 2976
QY 532 euValAspAsnThrTrpAsnThrHisIleAlaIleValGlyValAlaMetAla 552
Db 2977 TGTGTTGACACACACCTGGAACACACCCATATGCTACCGTTGGTAAGGCGCATGAGCT 3036
QY 552 erProGluGlyLysSerTyGluIleLeuLysGluHisAspValAspTyTrpValLeuValI 572
Db 3037 CACGCGAGGAGTACCTACCCATCTCCGCGAGCATGATGCTGATTACGTCGTGCTGG 3096
QY 572 lePheGlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleA 592
Db 3097 TGTTCGGTGTCTGCTAGTATTCTGCGGATGACATTAACAAATCTTATGATGCTGC 3156
QY 592 rgIleSerGluGlyIleTrpProGluGluIleLysGluArgTyPheTyThrAlaGluG 612
Db 3157 GTATCGCCGAGGATCTTGGCCCGATGAGGTAAAGCGGAGCTTCTTTACTGCGCG 3216
QY 612 lyGlyTyArgValAlaArgAlaSerGluThrMetArgAsnSerLeuLeu----- 629

Db 3217 GTGAATATCGTGTGCAAGATGAGCGAGCCCAACTATGCGCAACAGCTGATGATGAT 3276
QY 630 -----Tyr 630
Db 3277 TCCCTCTTCCCTCCCTTATCTACTGCGTGCATCAAGCTAATTCACAAAATCTAGGTAT 3336
QY 631 LysMetSerTyLysAspPheProGluLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 3337 AAAATGTCTTATTAACAATTTCAATCTCTTCCGTCGGCGCAAGCTGTCGACCGGTC 3396
QY 651 ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyTrpPheAspGluValPhe 670
Db 3397 CGTGGGTCAAAACTT-----CCACAGAAGGCCCTCAGCTCTCTACCTCGAAGAAGCTTTC 3453
QY 671 ThrSerGluAsnTrpMetValArgIleTyGlnLeuLysLysAspAlaGlnGlyArg 690
Db 3454 ACGAGCGAAGACTGGATCAATTCGATCTACAGGTCAAGGATCTTACAACTTGGCCGA 3513
QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLys 710
Db 3514 GACCACACACAGGCTGTTGCCCTTCGACAAA---GGTCTCAAGAAAAGCGGAGTACAAAG 3570
QY 711 Arg 711
Db 3571 AGG 3573
RESULT 10
ABT18969
ID ABT18969 standard; DNA; 1848 BP.
XX AC ABT18969;
XX DT 16-APR-2003 (first entry).
XX DE Aspergillus fumigatus essential gene #1327.
XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX OS Aspergillus fumigatus.
XX PN WO200286090-A2.
XX PD 31-OCT-2002.
XX PF 23-APR-2002; 2002WO-US013142.
XX PR 23-APR-2001; 2001US-0285697P.
XX PR 27-APR-2001; 2001US-0287086P.
XX PR 05-JUN-2001; 2001US-0295890P.
XX PR 09-JUL-2001; 2001US-0303899P.
XX PR 31-AUG-2001; 2001US-0316362P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX Disclosure; Page; 175pp; English.
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or

CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogs having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of *Aspergillus fumigatus* of the invention
 XX
 SQ Sequence 1848 BP; 363 A; 510 C; 447 G; 528 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1,01e-205 Length: 1848
 Score: 2090.00 Matches: 394
 Percent Similarity: 79.53% Conservative: 80
 Best Local Similarity: 66.11% Mismatches: 93
 Query Match: 55.57% Indels: 29
 DB: 8 Gaps: 5

US-10-028-384-6 (1-718) x ABT18969 (1-1848)

QY 63 PheTyrLysPheLeuAenTTPheAspAsp-----Arg-ThrTr 75
 DB 6 TTCTATAGCTTTGGATTGGTTGATGACCGCTAATCGCGTCCCAACATAGGACATG 65
 QY 75 pTyr-ProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrSerAl 95
 DB 66 GCATCCTCTGGACGCTGCACCGGTGGACAGTTATATATCCCGTCTCATGGTACGAGCGG 125
 QY 95 aPheLeuThPheAlaLeuAenTTPheLeuGlyLeuProLysLeuAspLysArgAsnValCy 115
 DB 126 CGTATCTACCATATCTTGGGA---TTCTTATATATCCCGTCCGATTCGCAACATCTG 182
 QY 115 sValLeuPheAlaProLeuPheSerGlyValThrAlaThrAlaThrTyrGluPheThrLy 135
 DB 183 CGTCTACTGGCGCCAGGATCTCCGCGCTGACTGCAATGGCAATGACTTCTGACATC 242
 QY 135 sGluLeuLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheLeuAlaLeuValPr 154
 DB 243 CGAGATGCTCTCTCGCCATCTGCAGGTCTCTTTGACGAGCTTTTCATGGATCGCCCC 302
 QY 154 oGlyTyrLeuSerArgSerValAlaGlySerTyrAspAsnGluAlaAlaAlaLeuLe 174
 DB 303 TGGTTACATCTCCCGATCAGTTGTGTGAAGCTACGATAACGAAGCGAATGGCATCTTCT 362
 QY 174 uLeuMetValThrPheMetPheTyrLysAlaGlnLysThrGlySerIleMetHisAl 194
 DB 363 GTTGTGTTCACATCTTCTATGGATCAAGGCTGTCAAAATGGGTCTATCATGTGGGG 422
 QY 194 aThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaThrGlyTyrValPheIl 214
 DB 423 AGCGCTGACCCGACATATCTACGGCTACATGGTGTGGCATGGGTGGGTATGTCTTCAT 482
 QY 214 eThrAsnLeuLeuProLeuHisValPheLeuLeuLeuLeuMetGlyArgTyrSerSery 234
 DB 483 TAGGACCTGATCCCGCTGCAGCTTTTGTCTCTGTCATGGGTAGATACAGCATCG 542
 QY 234 sLeuTyrSerAlaTyrThrThrTyrAlaAlaGlyThrValAlaSerMetGlnIlePr 254
 DB 543 CATCTACATATGATACCATGGTATGGCGTGGGACCTTTGGCTAGCATGCATGATCC 602
 QY 254 oPheValGlyPheLeuProLysArgSerAsnAspHisMetAlaAlaLeuGlyValPheGl 274

DB 603 CTTGTCGGATTTTGGCTATCTCCGAAACAGCAGCACCATGTCCGCTTGGGTGTCTTCGG 662
 QY 274 yLeuLeuGlnLeuValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysPh 294
 DB 663 CTGTGTTACGTTTGGCGCTTTCGGCAGTTTGTCCGAGCTTGTTCGAAGCAGGAGT 722
 QY 294 eLysValLeuMetMetValSerLeuPheLeuLeuValLeuValValGlyLeuSe 314
 DB 723 CCAGAGACTTCGACCGCCCATGATCTTCATCACCTCGGTCTGGTTTGGGTAGT 782
 QY 314 rAlaLeuThrTyrMetGlyLeuLeuAlaProTTPheGlyArgPheTyrSerLeuPhe 334
 DB 783 TTTTCTGACTGACGCGAGTATCGCTCTCTTGAGCGCGCGATCTTACTCTCTGGGA 842
 QY 334 pThrAsnTyrAlaLysIleHisIleProLysLeuAlaSerValSerGluHisGlnProVa 354
 DB 843 CACTGGCTATGCCAAATCCCATTCCTCATCTGCTCATCTGCGAACCAGCCAC 902
 QY 354 lSerTrpProAlaPhePheAspThrHisPheLeuLeuTyrLeuPheProAlaGlyVa 374
 DB 903 CGCTTGGCGAGGTTCTTCTTCGATCTGAACCTCTGATCTGGCTTTCGCGAGGTGT 962
 QY 374 lPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerValle 394
 DB 963 CTACATGTGCTTCCGTGACCTCAAGGACGACATGTCTCTGTCATTATCTACTCGGTCT 1022
 QY 394 uCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIleCy 414
 DB 1023 TCGGAGCTACTTCGCGGTGTATATGGTCCGACTAATGTGACTTACCCTTATTTGTG 1082
 QY 414 sValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu----- 429
 DB 1083 TGTTCGCGCTGCTTCGCGCTGTCTGCTCATCTCGACACCTATATATGCGGACTACCTCC 1142
 QY 430 -----AspPheLysThrSerAsp----- 435
 DB 1143 GACACAGCGCTGAAGCGAAACAGATGAAGACTCGCTCTTCAACCTCTTCGCTCAGT 1202
 QY 436 -ArgLysTyrAlaLysProAlaLeuLeuAlaLysLeuLeuValSerGlySerPh 455
 DB 1203 TCGGAAGCCCAATGTGGAATCACCTCCCATGTTTCTAAGATTATAGTACGCGGTCTGT 1262
 QY 455 eLlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSe 475
 DB 1263 TGTGCTACCTGCTCTCTGTTTGTTCGCACTGCACTGGGTACATCGAATGCATCTC 1322
 QY 475 rSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspAs 495
 DB 1323 TTCTCTTCGCTGTCTGCTGCTAGTGGATCGCTGACGGAAGCCAAATACATCATTCACGA 1382
 QY 495 pPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAl 515
 DB 1383 TTATCTGAGGCTTACTACTGCTTCGTCAGAACTACTCTCGAAACCCCAAAATCATGTC 1442
 QY 515 aTrpTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAs 535
 DB 1443 ATGTTGGGATTATGGGTATCAAAATCGGTGGCATCGCGACCGCCCAACCTTGGTTGACAA 1502
 QY 535 nAsnThrTrpAsnAsnThrHisIleAlaLeuValGlyValAlaMetAlaSerProGluGl 555
 DB 1503 CAACACTGGAAACAAACCCATATTTGCTACGTTGGTGAAGCGGATGAGTCAACGCGAGA 1562
 QY 555 uLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGl 575
 DB 1563 AGTCAGCTACCCCATCTCCGCGACATGATGTGATACGTGCTGGTGTGGTGTGGTGG 1622
 QY 575 yLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGl 595
 DB 1623 TCTGCTAGGTATTTCTGGCATGACATTTACAAATCTTATGGATGCTCGTATCGCGGA 1682
 QY 595 uGlyIleTTPProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrAr 615
 DB 1683 AGGTATCTGGCCCGATGAGTTAAAGAGCGGAGCTTCTTTACTGACGCGGTGAATATCG 1742

QY 615 gVallepAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyr 630
 Db 1743 TGTGACGATGGAGGACCCCAACATATCGCAACAGCTTGATGAT 1788

RESULT 11
 ABT18375
 ID ABT18375 standard; DNA; 1969 BP.
 AC ABT18375;
 XX
 XX 16-APR-2003 (first entry)
 XX
 XX Aspergillus fumigatus essential gene #733.
 DE
 XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.
 XX
 XX Aspergillus fumigatus.
 OS
 XX
 XX WO200285090-A2.
 PN
 XX
 XX 31-OCT-2002.
 XX
 XX 23-APR-2002; 2002WO-US013142.
 PF
 XX
 XX 23-APR-2001; 2001US-0285697P.
 PR
 XX 27-APR-2001; 2001US-0287066P.
 PR
 XX 05-JUN-2001; 2001US-0295890P.
 PR
 XX 09-JUL-2001; 2001US-0303899P.
 PR
 XX 31-AUG-2001; 2001US-0316362P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 PI
 XX WPI; 2003-093124/08.
 DR
 XX
 XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.
 XX
 XX Disclosure; Page; 175pp; English.

The invention relates to novel purified or isolated nucleic acids of essential genes of Aspergillus fumigatus. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as A. fumigatus, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by A. fumigatus, or to prevent or inhibit formation on a surface of a biofilm comprising A. fumigatus. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of A. fumigatus to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organisms invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This polynucleotide sequence represents one of the essential genes of Aspergillus fumigatus of the invention

Sequence 1969 BP; 389 A; 540 C; 480 G; 560 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.03e-200 Length: 1969
 Score: 2042.00 Matches: 394
 Percent Similarity: 74.41% Conservatives: 80
 Best Local Similarity: 61.85% Mismatches: 93
 Query Match: 54.29% Indels: 70
 DB: 8 Gaps: 6

US-10-028-384-6 (1-718) x ABT18375 (1-1969)

QY 63 PheTyrLysPheLeuAsnTrpPheAspAsp-----72
 Db 6 TTCTATAGCTTTTGGGATTTGGTTTGGTATGATGACCGTATGTTCCCGCCACCAAGAGTGC 65
 QY 73 -----Asg 73
 Db 66 ATGAAGCAACAGGTCACTGTCTGGAGTGTCTTAGGCTAATCGGCGTCCCAACATAGG 125
 QY 74 ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrTh 93
 Db 126 AACATGGCATCTCTGGGAGTGTACCGGTGGCAGCTTATATCCCGGTCTCTATGGTGAC 185
 QY 93 rSerAlaPheIleThrPheHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAs 113
 Db 186 GAGCGCGTGTATCATCATATCTTGGCA---TTCTTACTATCCCGTGTGATATTCGCA 242
 QY 113 nValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPh 133
 Db 243 CATCTGCGTCTCTACTGGCGCAGGATTCTCCGCGCTGATTCGCAATGTACTTGCT 302
 QY 133 eThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaI 152
 Db 303 GACATCCGAGATGTCCTTCGCCCATCTGCAGGTCTCTTCGACGAGCTTCATCGGAAT 362
 QY 152 evalProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaI 172
 Db 363 CGCCCTGGTTACATCTCCCGATCAGTTGCTGGAAGCTACGATTAACGAACGATTCGCAT 422
 QY 172 eThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMe 192
 Db 423 CTTCCTGCTTGTGTTCACATCTTTCTATGATCAAGCTGTCAAAATGGGTCTATCAT 482
 QY 192 tHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyGlyTyrVa 212
 Db 483 GTGGGGAGCGGTGACCGCATCTTTCTACGGCTACATGGTGTGCGCATGGGTGGGTATGT 542
 QY 212 lPheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSe 232
 Db 543 CTTCATTAGCAACCTGATCCCGCTGCACGTTTTGTTCCTTCTGTGATGGGTAGATACAG 602
 QY 232 rSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGl 252
 Db 603 CACTGCGATCTACATTTAGCTATACCATGGTATGCGTGGGACTTTGGCTAGCATGCA 662
 QY 252 nIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeu---- 270
 Db 663 GATTCCTCTGTCGATTTTGGCTATCCGAACAGCGACCATGTCCGCTTGGGTAG 722
 QY 271 -----GlyValPheG 274
 Db 723 GTCTGTGTTTCGCAAGTTCAAATGGCGGTTCAGGAGCTGACATTTTGGAGGTGCTTCG 782
 QY 274 lYLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysP 294
 Db 783 GCCTGCTTACGTTGTGGCTTCGCGAGTTTTCGCGAGCTTCGTTCCCAAGCAACGACT 842
 QY 294 heLysValIleMetMetValSerLeuPheLeuIleValLeuGlyValValGlyLeus 314
 Db 843 TCCAGAGACTTCTGACCGCATCATCTCATCACTTCGGTCTCGGTTTGGTGGCTAG 902
 QY 314 erAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpA 334
 Db 903 TTGTTCTGACTGTGACCGGAGTGTGCTCTTTGGAGCGCCGATCTTACTCTCTTGGG 962

97 eTrpHisAlaLeuArgAsnTrpLeuGlyLeuProPheAspLeuValCysValle 117
117 uPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyr----- 131
639 CTTGACCTGGATTCTGGATTGACTGCGTGGCGGACTTATCTGTCAGTGAACAAT 698
132 -----GluPheThrLysGluLeuLysAs 139
699 TATCTCATCTGATACACTAATGAGAGGTTGTAGTTTCCACCTGAAATGTCAT 758
139 pAlaSerAlaGlyLeuLeuAlaGlyPheLeuAlaLeuValProGlyThrLysSer 159
759 ACCATGAGCTGGTCTATTGGCGCGCTTTTATGGCATTTGATCCCGGATACATCTCG 818
159 gSerValAlaGlySerTyrAspAsnGluAlaLeuAlaLeuThrLeuLeuMetValThrPh 179
819 ATCTGTCGCGGTTCTTATGACAACGAAGCCATTGCCATCTCTCTGTGAGCTCCCT 878
179 eMetPheTrpLeuLysAlaGlnLysThrGlySerLeuMetHisAlaThrCysAlaAlaLe 199
879 CTACTCTTGATTAAGCGCGTCAAAACCGTAGCTCATTTTGGGGGTATGATCTGCTT 938
199 uPhefYrPhefYrMetValSerAlaTrpGlyGlyYrValPheLeuThrAsnLeu---- 217
939 GTTCTACGGGTGGTGTGCTGCATGGGTGGTGTACGTTTTCATCCCAACAG-TATGT 997
218 -----IleProLeuHisValPh 223
998 CGTCCGCGCCTCAATTGATTCTGTTTACTCTTTTCAGTGAATTCATGACGCTT 1057
223 eLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTrpThrTrpTyr 243
1058 TGTTCATTTGTCATGGCGAGGTTCAAAACCGGCTTTATACCGCTTACTCTCTGCTGA 1117
243 rAlaLeuGlyThrValAlaSerMetGlnLeuProPheValGlyPheLeuProLeuArgse 263
1118 TGTCAATGAACTATGCTCCATGAGGTCCTTTGAGGTTCTCCCATCCGAC 1177
263 rAsnAspHisMetAlaLeuGlyValPheGlyLeuLeuLeuAlaPheGlyAs 283
1178 CTCTGAGCACATGGCGGCTGGGTGTTTGGGTCTGTACAGCTGATCGGATTCGTCGA 1237
283 pPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPh 303
1238 AGTGTCCGACGACTGCTGCTGCGAGCAATTCAGCTCTCTCAAGCTTTTGTCT 1297
303 eLeuLeuLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuLeuAl 323
1298 GCGCGTATTCTGCTCAGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1357
323 aProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle----- 340
1358 CCGCTTGGTGGAGATTATCTCTTTGGATGACTGCTATGCGAGGTCACAGTGA 1417
341 -----His-IleProLeuLeuAla 347
1418 GTCAATGCCATACCTTCGGGATCTATATTATTGTCAACATAGTCCCATATTGCTT 1477
347 eValSerGluHisGlnProValSerTrpAlaPhePhePheAspThrHisPheLeu 367
1478 CCGTCTCCGAACACCGCCACCGCTTGGCGCTCTTCTTCTGCTGCTGCTGCTGCTGCT 1537
367 leTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLysValP 387
1538 TCTCTTTTTCCTGCGGTGCTCTTCTGGTGTTCAGGAGCTTCGGGATGAGCAGATCT 1597
387 heValIleAlaTrpSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetL 407
1598 TCATCATCTTATGCGGTCTCAGTGCCTATTTTGGCGGTGCTATGCTGCTGCTTATGC 1657
407 euThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspI 427

1658 TTGTCTACCGCTGTGTGTCTGTGTCTCGCCATTGCGTCTCTCCAACTCTCTCGAGG 1717
427 leTyrLeuAsp----- 430
1718 CGTATATTGACCCCGCTCATCCCGAAGGACGAGGAAGCTGGGAGTCTCAGACGCGAG 1777
431 -----PheLysThrSerAspArgLysTyrAlaLeuLysProAla----- 443
1778 TTGTCTCAAGTCCAAAGCGGCAAGATGCGCTGCCAACGCCAATAAGAGCGGGTCT 1837
444 -----AlaLeuLeuAlaLysLeuIleValSerGlySerPhe----- 455
1838 CTTTCACAGGTATTTGAGCGCAAGTCTGTCTCCGGCATCTTTGGTCTCGACACTCGAT 1897
456 -----IlePheTyrLeuTyrLeuPheValPheHisSerThrTrpV 469
1898 TTGTGTGGTTCCT 1957
469 alThrArgThrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyL 489
1958 TGATCTCAACAGCGTATCTTCCCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2017
489 ysLeuAlaLeuLeuAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspG 509
2018 GCCAAATATCATTTGATGATTTCCGAGAGCTTACTACTGATTCGCCAAACACCGCG 2077
509 luAspSerLysValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGlyMetAlaAspA 529
2078 AAGACAGGCTCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2137
529 rgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaLeuValGlyLysA 549
2138 GCCCACCTCTGTTGATTAACAATACCTGGAATAACCCACCATTCGCCAGTTGTGAAG 2197
549 laMetAlaSerProGluGluLysSerTyrGlyLeuLeuLysGluHisAspValAspTyrV 569
2198 CCATGGCTTCCACGAAGATGTCATATCTTCTGAGGAAGCATGATGCTGATTCAG 2257
569 alLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeu 589
2258 TTCCTGTGATCTTTGGGGGCTTATTTGGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2317
589 rpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyr 609
2318 GGATGTTAGATCTCAAGGTGATGCTGCTGAGGCTGAGGAGTCAACTACTTCTA 2377
609 hrAlaGluGlyGlyTyrArgValAspAlaArgAlaSerGlu----- 622
2378 CTCAAAGAGGGGAGTATGCTGCTGATGACAGGCGCTG-CGTCTATTTTGTATTATTGCC 2436
623 -----ThrMetArgAsnSerLeuLeuTyrL 631
2437 TGGCGAAATTTGCTGCAATCTGGCGAGCACCCCTACTATGAAGAACTCTCTCATGTACA 2496
631 ysMetSerTyrLys----- 635
2497 AATGTCTTACTA-CCGGTAGGTCAAAATACATTTGCTGATAGTAACTCTCTGACAT 2555
636 -----AspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnM 654
2556 ATCTCAGCTTCCCGAGCTTATGTTGGACACCCCGCTCAAGACAGGTTTCGAGGCCAAA 2615
654 etIleThrProLeuAspValProLeuAsp-TyrPhe----- 666
2616 TTATCCCTCTCAAGTGT-----ACTCTGATACTCTTGTAGTTTACGGTTTGAAGAT 2672
667 -----AspGluValPheThrSerGluAsn 674
2673 AAGTGTATAGACGCTAATGCTGCTCATTTCTCTGACAGGAGCGTTCACTCCGAAAT 2732
675 TrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAsp 694

Db 2733 TGGATCGTCAGATCTACAGGTCAGAGAGAGATCCCATTTGGACGAGACACCAAGGCC 2792

QY 695 Val-----GlyGluLeuThrArgSerSerThrLysThrArgSerIle 709

Db 2793 GTTACTGCTGCGACGGGGTAAGAAGTTGAAGAAGAGTCTAGTCCAGTGAGGGCGTG 2852

QY 710 LysArg 711

Db 2853 AAGCGG 2858

RESULT 13

ID ADB69178 standard; DNA; 4738 BP.

XX ADB69178;

XX DT 04-DEC-2003 (first entry)

XX C. neoformans genomic DNA sequence SEQ ID NO:305.

XX ds; gene; fungicide; gene therapy; infection.

XX Cryptococcus neoformans.

XX PN WO2003052076-A2.

XX PD 26-JUN-2003.

XX PF 17-DEC-2002; 2002WO-US040225.

XX PR 17-DEC-2001; 2001US-0341261P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Zamudio C, Eroshkin AM;

XX DR WPI; 2003-533017/50.

XX DR P-PSDB; ADB70261.

XX New nucleic acid, useful for preparing a composition for treating an infection caused by Cryptococcus neoformans.

XX Claim 3; SEQ ID NO 305; 136pp; English.

CC The invention relates to a novel purified or isolated Cryptococcus neoformans nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by Cryptococcus neoformans.

CC The present sequence represents a C. neoformans sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX SQ Sequence 4738 BP; 1162 A; 1160 C; 1070 G; 1346 T; 0 U; 0 Other;

Alignment Scores:

| Pred. No.: | Length: | Score: |
|------------------------|---------|---------|
| 3,59e-199 | 4738 | 2032.50 |
| Matches: | 433 | |
| Percent Similarity: | 60.86% | |
| Best Local Similarity: | 48.26% | |
| Query Match: | 159 | |
| Indels: | 191 | |
| Gaps: | 14 | |

US-10-028-384-6 (1-718) x ADB69178 (1-4738)

QY 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaIleSerSerArg 32

Db 1079 GAAAGCTGTGGCGCTTCATCATCTCGCGTTTATATGCGGTGCTCGATTGGAAGTGA 1138

QY 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPhe----- 46

Db 1139 TTGTTTGGCGTGATCAGATTCCGAATCTGTCTCATCCACGAATTTGTAGTCAAAATCTATCCG 1198

QY 47 -----AspProThrPheAsnTyr-- 52

Db 1199 AGGANTCTCTCAATGCTAAGCTGTGTCTTTCATTTCAGTACGCCCTGGTTCACCTA-GT 1257

QY 53 -----ArgAlaThrLysTyr 57

Db 1258 AAGGCTATTTCTTCGAGGTCAAACTTCCTACTAGTATTTATTCCTAGCGAGCCCTCGAAGT 1317

QY 57 rLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAspArgThrTrpTrp 77

Db 1318 TCTTGTGTTAACAGAGGGTTTCTACGAGTCTTGGAATGCTTTGACCCCTCCGCTGGTACCC 1377

QY 77 oLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPhe11 97

Db 1378 TCTGGCAGACTGTCGTCACAGCTCTATCTGGCTTGATGGTCAGCTCGACTGAT 1437

QY 97 eTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysVal1e 117

Db 1438 TTGGCATGCTCTTCGGCA--ATCAATATGCCCGTGGACATTCGCAATGTCTGTCTCT 1494

QY 117 uPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyr----- 131

Db 1495 CTTGCACCTGGATTTTCTGGATTGACTGCTGGCGGACTTATCTGTCAGTGAACAAAT 1554

QY 132 -----GluPheThrLysGluLeuLysAs 139

Db 1555 TATCTGATCATTCATACACATAACATGAGACGCTGTAGTTTACCACCTGAATGCTAC 1614

QY 139 pAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerAr 159

Db 1615 ACCATCAGCTGCTCTATTGGCGCGCTTTCATTGGCATGTACCCGGATACATCTCTCG 1674

QY 159 gSerValIaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPh 179

Db 1675 ATCTGTGCGCGGTTCTTATGACAAGAGCAATTCATCTCTCTTGTAGAGTCTCT 1734

QY 179 eMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAla1e 199

Db 1735 CTACTCTTGATTAGGCGCTCAAAACCGGTAGTCTATTTGGGGTATGATCACTGCCCT 1794

QY 199 uPheTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleThrAsnLeu----- 217

Db 1795 GTTCTACGGTGGATGGTGTCTGTCATGGGGTGTACGTTTTCATCACCACAG-TATGT 1853

QY 218 -----IleProLeuHisValPh 223

Db 1854 CGTCGGCGCTCAATTGAATGTCTGTTTACTCTTTTGCAGTGTATTCATTCAGCCCTT 1913

QY 223 eLeuLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThrThrTrpTy 243

Db 1914 TGTCTCATGTCATGGCAGGTTCACACACCGGCTTATACCGCTTACTCTCTCCGGTA 1973

QY 243 rAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArgSe 263

Db 1974 TGTCAATGGAATATCGCTCCATCGAGTCCCTTTGTGGAGTCTCTCCCATCCGAAC 2033

QY 263 rAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAs 283

Db 2034 CTCTGAGCACATGGCGGCTTGGGTGTTTTCGGTCTTGTACAGCTGATCGATTCGCGA 2093

QY 283 pPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetValSerLeuPh 303

Db 2094 AGTCGTCCGAGACTCGTGGCTGGCAAGCAATTCAGCTCTCTCTCAAGCTTTTTCGT 2153

QY 303 eLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAl 323

Db 2154 GGCGGTATTCGCTCAGTTTGTCTGCCCTCGTCACTTGCATCTCTCTGATGATGATCGC 2213

QY 323 aProThrPheGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaIleValle----- 340

Db 2214 CCCCCCTCGCGAAGATTTTATCTCTTTGGGATACCTGGCTATGCGAAGCTCCACAGTGA 2273

341 QY -----His-IleProlleleAlas 347
 2274 Db GTCAATGCCATACCTTCGGGATCTATATTATTGTCAACATAGTGCCTATTATTCCT 2333
 347 QY erValSerGluHisGlnProValSerTTPProAlaPhePheAspThrHisPheLeuI 367
 2334 Db CCGTCTCCGACACACGACCCACCGCTTGGCCCTCATCTACTTTGACCTCGAATGCTTA 2393
 367 QY letPLeuPheProAlaGlyValPheLeuLeuPheLeuAspLeuLeuHisValP 387
 2394 Db TCTCTTTTTCCTTCGGGCTGCTCTCTGGTGTTCGAGGAGCTTCGGGATGAGCAGATCT 2453
 387 QY heValIleAlaTyrSerValLeuCysserTyrPheAlaGlyValMetValArgLeuMetL 407
 2454 Db TCATCATCATTTATGCGGTTTCAGTGCCATTTTTCGGGCTGTCATGTTTCGACTTATGC 2513
 407 QY eutThrLeuThrProValIleCyssValSerAlaAlaValAlaLeuSerIysIlePheAspI 427
 2514 Db TTGTCATCAGCGCTGTGTCTGTGTTCCTCGCCATTCGCTTCTCCAAACTTCTCGAGG 2573
 427 QY letTyrLeuAsp----- 430
 2574 Db CGTATATTGACCCCGTCATCCCGAAAGCGACGAGAGCTGGCGAGTCTCAGACGCGAG 2633
 431 QY -----PheLysThrSerAspAlyLysTyrAlaIleLysProAla----- 443
 2634 Db TTGTCCTCAGTCCAAAGCGAGAGAGATGCGCGCTGCCAAGCCCAATAAGAGCGGGTCT 2693
 444 QY -----AlaLeuLeuAlaLysLeuIleValSerGlySerPhe----- 455
 2694 Db CTTTCACAGGTATTTTGGCGGCAAGTCTGCTCCGCGCATCTTTGGTCTCGACACTCGAT 2753
 456 QY -----IlePheTyrLeuTyrLeuPheValPheHisSerThrTrp 469
 2754 Db TTGCTGTGGTTCCATCTCTGCTCTTCCTCTTCATCTTTGCTCTCTGTCGACATATG 2813
 469 QY alThrArgThrAlaTyrSerProSerValValLeuProSerGlnThrProAspGlyL 489
 2814 Db TGACTTCAACAGCGTATCTTCGCGCTTCAGTGGTACTTGCATCGCAACCCCGATGGTA 2873
 489 QY ysLeuAlaLeuIleAspAspPheArgGluAlaTyrTyrLeuPheMetAsnSerAspG 509
 2874 Db GCCAAATATCATGTATGATTTCCGAGAGGCTTACTACTGATTCGCCAAACACCCCG 2933
 509 QY luAspSerLysValAlaAlaTyrTyrAspTyrGlyTyrGlnIleGlyMetAlaAspA 529
 2934 Db AAGACAGCGCTCATGCTCTGCTGGATACGGCTACCATCGCTGGTATGCTGCTGATC 2993
 529 QY igThrThrLeuValAspAsnAsnThrTyrAsnAsnThrHisIleAlaIleValGlyLysA 549
 3054 Db GCCCCACCTTGTGTATACATATCTGGATTAACACCCACATATGCCACAGTTGGTAAAG 3053
 549 QY laMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrV 569
 3054 Db CCATGGCTTCCACGAAGATGTCGATATCTCTATCTTGAGGAAGCATGATGCTGATTAAC 3113
 569 QY alLeuValIlePheGlyGlyLeuIleGlyPheGlyAspAspIleAsnLysPheLeuT 589
 3114 Db TTCTTGTGATCTTGGGGCTTATTTGGGCTACTCTGGTGACGATATCAACAGTTTGT 3173
 589 QY rpMetIleArgIleSerGluGlyIleTyrProGluGluIleLysGluArgTyrPheTyrT 609
 3174 Db GSATGTTAGGATCTCACAAGGTGAATGGCCCTGACGAGGTGCGAAGTCAACTACTTTA 3233
 609 QY hrAlaGluGlyGluTyrArgValAspAlaArgAlaSerGlu----- 622
 3234 Db CTCAAGAGGGAGTATGCTGTCGATGACAGGGCGTG-CGTCATTATTGTTATTATGTC 3292
 623 QY -----ThrMetArgAsnSerLeuLeuTyrL 631
 3293 Db TGGCGAAATTGCTGACAAATCTGGGACGACCCCTACTATGAAGAACTCTCTCATGTACA 3352

631 QY ysMetSerTyrLys----- 635
 3353 Db AAA-GTCTTACTA-CCGGTAGGTCAAAATACATTGTGATAGATGAACATCGTCTGACAT 3411
 636 QY -----AspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnM 654
 3412 Db ATCTCAGCTTCCCGAGCTTTATGGTGACACCCCGCTCAAGACAGGGTTCGAGGCCAAA 3471
 654 QY etIleThrProLeuAspValProLeuAsp-Tyr-Phe----- 666
 3472 Db TTATCCCTCTTAAACAGTGT----ACTCTTGATACTCTTGGTAGTATTAGGTTTGAAGAT 3528
 667 QY -----AspGluValPheThrSerGluAsn 674
 3529 Db AAGTGTATAAGACGCTAATGTGTCACATTCCTCGACAGCAAGCGTTCATCCGAAAT 3588
 675 QY TrpMetValArgIleTyrGlnLeuLysLeuLysAspAlaGlnClyArgThrLeuArgasp 694
 3589 Db TGGATCGTCAGATCTACAAGGTCAAGAGAGATCCCATTTGGAGGAGACACACAGGCC 3648
 695 QY Val-----GlyGluLeuThrArgSerSerThrIlysthrArgSerIle 709
 3649 Db GTTACTGCTGGAACGGGGTAAAGATTGAAGAGAGTCTTAGTCCAGTGGGGCGTG 3708
 710 QY LysArg 711
 3709 Db AAGCG 3714
 RESULT 14
 ADD94789
 ID ADD94789 standard; DNA; 2417 BP.
 XX
 AC ADD94789;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Drosophila melanogaster STT3 gene sequence.
 XX
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cystic; immunosuppressive; antitumor therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide;
 XX
 OS fruit fly; gene; ds; STT3.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO2003054008-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 PR 20-DEC-2001; 2001US-00028384.
 XX
 PA (COMP-) COMPATIGENE INC.
 XX
 PI Perreault C, McBride X;
 XX
 DR WPI; 2003-559122/52.
 XX
 DR P-PSDB; ADD94790.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 XX Claim 6; SEQ ID NO 7; 66bp; English.
 XX
 XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-

CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, pancreatic cancer, testicular cancer, breast
 CC cancer, melanomas, sarcomas, cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the *Drosophila* melanogaster *Sirt3* gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4,4e-199 Length: 2417
 Score: 2027.50 Matches: 394
 Percent Similarity: 68.22% Conservative: 117
 Best Local Similarity: 52.60% Mismatches: 187
 Query Match: 53.91% Indels: 51
 DB: 10 Gaps: 7

US-10-028-384-6 (1-718) x ADD94789 (1-2417)

QY 12 PheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSer 31
 DB 113 TACAGAGCGCTATACCTTCCGATCTGCTAATCGCTGCGTGGCGGATTTCCCT 172
 QY 32 ArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsn 51
 DB 173 CGCCTCTTCCGCGTCATCCGTTCCGAGTTCGATATTCATCCATGAGTTTGCATCC 232
 QY 52 TyrArgAlaThrLysLeuValAlaAsnSerPheTyrLysPheLeuAsnTrpPheAsp 71
 DB 233 TACCGGGCCACCGCTATCGTGCAGATGGTGGTACAACTTCTCACTCGTTCGAC 292
 QY 72 AspArgThrTrpTyrProLeuGlyArgValThrGlyLysThrLeuTyrProGlyLeuMet 91
 DB 293 GAGCGCGATGATATCCCGCTCGGAGGATTTGGCGGTACCGCTATCCCGGCTGATG 352
 QY 92 ThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIle 111
 DB 353 ATTACGTCGGCGGATC---CATTTGCTGCTGCACGCTACTCAACATACCGTCCATAT 409
 QY 112 ArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTrp 131
 DB 410 CGTGACATCTGGTGTCTTCCGCGCGGATCTTCAGTGGCTGACCTCCATCCACCTAC 469
 QY 132 GluPheThrLysGluIleLysAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAla 151
 DB 470 CTGCTGACCAAGAGCTGTGTGCGCGCGCGCTCTTCGCGCGAGCTTCATCGCC 529
 QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 DB 530 ATCGTGGCTGGCTATCAGTAGGTGCGGTGGCTGGATCGTACGATACAGGGGCGATGCC 589
 QY 172 IleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIle 191
 DB 590 ATATTGCGGCTGCAGTTCACCTACTCTCTGTGGGTGGCTCAGTGAAGAGCTGGATCCG 649
 QY 192 MetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyr 211
 DB 650 TTCTGTGGCGCGAGCGCTTTGTCTTCTTACATGTTGCTCCCGCTGGGTGGCTAC 709
 QY 212 ValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTyr 231
 DB 710 GTGTTTCATCATCACTGATACCGCTGCACGCTCTCTGCTACTGCTATTATGGCGAGTAC 769

QY 232 SerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMet 251
 DB 770 TCGCGCGCTGCTGCTACCCAGCTTACAGACCTTACATCTCTGGAGCTGTCTCTCATG 829
 QY 252 GlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGly 271
 DB 830 CAGATCCCTTCTGTTGGATTTCACCGATACGACAGTGAACACATGGCTGGCTGGGA 889
 QY 272 ValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThr 291
 DB 890 GTGTTTGTCTCTTATGCGCTGCGCACCTTGGCGCATTTGCGAGTCCGCTGCTGGCGC 949
 QY 292 AlaLysPheValIleMetMetValSerLeuPheLeuIleLeuValValVal 311
 DB 950 AAGAGTTCGGAAGCTTTCATGTCGCGGATTTGCTGGGCGCTGGCGCTTTGTG 1009
 QY 312 GlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSer 331
 DB 1010 GCGTGTGCTGCTCACCATCTGCGGCTTGTGCGCGCTGCGAGTGCAGCTTCTACTCG 1069
 QY 332 LeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHis 351
 DB 1070 CTGTGGATCTGGCTACGCCAGATCCACATTCCTCATCTGCTGCGAGCAT 1129
 QY 352 GlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPhePro 371
 DB 1130 CAGCCACACCTTGTCTCTGTTCTTGTATCTGCACATCTGCTGGTGGCTTCCCA 1189
 QY 372 AlaGlyValPheLeuLeuPheLeuAspGluHisValPheValIleAlaTyr 391
 DB 1190 GTGGAGTGTGTACTGCATCAAGCAGATCAACGACGCGCTTTTCGTGGTGTGTAC 1249
 QY 392 SerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrPro 411
 DB 1250 GCATCAGTGGCGTTTACTTCTGCTGGTGTGTGTGCTGGTGTGTGATGACCTTCACGCG 1309
 QY 412 ValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeuAspPhe 431
 DB 1310 GTGGTGTGATCTGCTGCGGAGTGGCTTTTCGGGACTGTGATGTTGTTCTTCAAGAG 1369
 QY 432 LysThrSerAspArg-----LysTyrAlaIleLys 441
 DB 1370 GATTCGTCTAAGCAAGTGGCGACAGCTGACCAAGCTGCGCAAGCTGAAGCATCTCTAAG 1489
 QY 442 ProAlaAlaLeuLeu-----AlaLysLeuIleValSerGlySer 454
 DB 1490 CATGATGCCACAGAGTACTGGGTGCTGCTCAACCTCAAGAGTATTGTATTATTTGGCC 1549
 QY 455 PheIlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyr 474
 DB 1550 GTTCTAATGCTCTGATGATGTTGCTGCTCCACTGCGAGTGGGTGACAGCAATGCTCTAC 1609
 QY 475 SerSerProSerValValLeuPro---SerdGlnThrProAspGlyLysLeuAlaLeuIle 493
 DB 1610 TCCAGTCCCTTCCATTTGTTGCTTCCACACAGTCAAGATGGATCCCGACATTTTA 1669
 QY 494 AspAspPheArgGluAlaTyrTyrTrpLeuArgMetIleSerAspGluAspSerLysVal 513
 DB 1670 GACGATTTTCAGAGAGCTTACTTCTGCTTTCGAGAGAACCTGCGCATGATGCTCGCT 1729
 QY 514 AlaAlaTrpTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuVal 533
 DB 1730 ATGCTCTGTGGATTCAGGATACCGATACCGAATGGCAACACAGACAGCGTAGTG 1789
 QY 534 AspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro 553
 DB 1790 GATAATAATACGTGAACAATAGTACATAGCTGCTGGTGGCAAGGCAATGTCTTCAACC 1849
 QY 554 GluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePhe 573

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:57:05 ; Search time 139.657 Seconds
(without alignments)
3654.277 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761
Sequence: 1 MGSDRSCLVSVFQTLKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh

-MODE=frame+ p2n model -DEV=xlh
-Q/cgn2_1/USFTO_spool/US10028384/runat 14122004 131520 8052/app_query.fasta_1.3740
-DB=Issued Patents NA -OFMR=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.1.259 @runat 14122004 131520 8052 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXI=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|---------------------|
| 1 | 3751 | 99.7 | 2157 | US-09-614-221A-318 | Sequence 318, App |
| 2 | 954.5 | 25.4 | 867 | US-09-248-796A-3089 | Sequence 3089, App |
| 3 | 892 | 23.7 | 1386 | US-09-270-767-11648 | Sequence 11648, A |
| 4 | 891.5 | 23.7 | 900 | US-09-270-767-12331 | Sequence 12331, A |
| 5 | 711 | 18.9 | 594 | US-09-248-796A-3090 | Sequence 3090, App |
| 6 | 648 | 17.2 | 1660 | US-09-270-767-15440 | Sequence 15440, App |
| 7 | 648 | 17.2 | 1660 | US-09-270-767-15440 | Sequence 15440, A |
| 8 | 595 | 15.8 | 507 | US-09-513-999C-1965 | Sequence 1965, App |
| 9 | 591.5 | 15.7 | 914 | US-09-270-767-12856 | Sequence 12856, A |
| 10 | 509 | 13.5 | 1660 | US-09-270-767-15440 | Sequence 15440, App |
| 11 | 509 | 13.5 | 1660 | US-09-270-767-15440 | Sequence 15440, A |
| 12 | 461.5 | 12.3 | 503 | US-09-270-767-27262 | Sequence 27262, A |

| | | | | | | |
|----|-------|------|-----|---|---------------------|--------------------|
| 13 | 450 | 12.0 | 616 | 3 | US-09-328-111-332 | Sequence 332, App |
| 14 | 409 | 10.9 | 487 | 3 | US-09-385-982-213 | Sequence 213, App |
| 15 | 396 | 10.5 | 487 | 4 | US-09-702-705-1655 | Sequence 1655, App |
| 16 | 396 | 10.5 | 487 | 4 | US-09-736-457-1655 | Sequence 1655, App |
| 17 | 396 | 10.5 | 487 | 4 | US-09-614-124B-1655 | Sequence 1655, App |
| 18 | 396 | 10.5 | 487 | 4 | US-09-671-325-1655 | Sequence 1655, App |
| 19 | 396 | 10.5 | 487 | 4 | US-09-658-824-1655 | Sequence 1655, App |
| 20 | 389 | 10.3 | 407 | 4 | US-09-702-705-375 | Sequence 375, App |
| 21 | 389 | 10.3 | 407 | 4 | US-09-702-705-1271 | Sequence 1271, App |
| 22 | 389 | 10.3 | 407 | 4 | US-09-736-457-375 | Sequence 375, App |
| 23 | 389 | 10.3 | 407 | 4 | US-09-614-124B-375 | Sequence 1271, App |
| 24 | 389 | 10.3 | 407 | 4 | US-09-614-124B-375 | Sequence 375, App |
| 25 | 389 | 10.3 | 407 | 4 | US-09-614-124B-1271 | Sequence 1271, App |
| 26 | 389 | 10.3 | 407 | 4 | US-09-671-325-375 | Sequence 375, App |
| 27 | 389 | 10.3 | 407 | 4 | US-09-671-325-1271 | Sequence 1271, App |
| 28 | 389 | 10.3 | 407 | 4 | US-09-589-184-375 | Sequence 375, App |
| 29 | 389 | 10.3 | 407 | 4 | US-09-589-184-375 | Sequence 375, App |
| 30 | 389 | 10.3 | 407 | 4 | US-09-658-824-1271 | Sequence 1271, App |
| 31 | 383 | 10.2 | 407 | 4 | US-09-702-705-588 | Sequence 588, App |
| 32 | 383 | 10.2 | 407 | 4 | US-09-736-457-588 | Sequence 588, App |
| 33 | 383 | 10.2 | 407 | 4 | US-09-614-124B-588 | Sequence 588, App |
| 34 | 383 | 10.2 | 407 | 4 | US-09-671-325-588 | Sequence 588, App |
| 35 | 383 | 10.2 | 407 | 4 | US-09-589-184-588 | Sequence 588, App |
| 36 | 383 | 10.2 | 407 | 4 | US-09-589-184-588 | Sequence 588, App |
| 37 | 324.5 | 8.6 | 560 | 4 | US-03-270-767-1318 | Sequence 1318, App |
| 38 | 324.5 | 8.6 | 560 | 4 | US-03-270-767-16600 | Sequence 16600, A |
| 39 | 324 | 8.6 | 292 | 4 | US-09-313-294A-4442 | Sequence 4442, App |
| 40 | 317 | 8.4 | 414 | 4 | US-09-513-999C-271 | Sequence 271, App |
| 41 | 278.5 | 7.4 | 269 | 4 | US-09-313-294A-906 | Sequence 906, App |
| 42 | 269.5 | 7.2 | 265 | 4 | US-09-313-294A-910 | Sequence 910, App |
| 43 | 268 | 7.1 | 291 | 4 | US-09-313-294A-4834 | Sequence 4834, App |
| 44 | 260 | 6.9 | 250 | 4 | US-09-513-999C-1438 | Sequence 1438, App |
| 45 | 257 | 6.8 | 302 | 4 | US-09-313-294A-6869 | Sequence 6869, App |

ALIGNMENTS

RESULT 1
US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; CURRENT FILING DATE: 2000-07-12
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-614-221A-318

Alignment Scores:
Pred. No.: 0
Score: 3751.00
Percent Similarity: 99.86%
Best Local Similarity: 99.86%
Query Match: 99.73%
DB: 4
Length: 2157
Matches: 717
Conservative: 0
Mismatches: 1
Indels: 0
Gaps: 0

US-10-028-384-6 (1-718) x US-09-614-221A-318 (1-2157)

QY 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleuLeuValVal 20
Db 1 ATGGGATCCGACCGGTCGTGTGTTTGTCTGTGTTTCAGACCATCTCAAGCTCGTCATC 60

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 107196.132
;; CURRENT APPLICATION NUMBER: US/09/248.796A
;; CURRENT FILING DATE: 1999-02-12
;; PRIOR APPLICATION NUMBER: US 60/074,725
;; PRIOR FILING DATE: 1998-02-13
;; PRIOR APPLICATION NUMBER: US 60/096,409
;; PRIOR FILING DATE: 1998-08-13
;; NUMBER OF SEQ ID NOS: 28208
;; SEQ ID NO 3089
;; LENGTH: 867
;; TYPE: DNA
;; ORGANISM: Candida albicans
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (25)
;; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

Alignment Scores:
Pred. No.: 4.16e-106 Length: 867
Score: 954.50 Matches: 178
Percent Similarity: 77.82% Conservative: 36
Best Local Similarity: 64.73% Mismatches: 58
Query Match: 25.38% Indels: 3
DB: 4 Gaps: 1

US-10-028-384-6 (1-718) x US-09-248-796A-3089 (1-867)

QY 446 LeuAlaLysLeuValSerGlySerPheIlePheThrLeuValPheHis 465
DB 37 TTGTCAGAGTTGGTTTACTGACATTTACATTTTACCTTTTCTACITTTGTTTACAT 96
QY 466 SerThrTrpValThrAlaThrSerSerProSerValValLeuProSerGlnThr 485
DB 97 TGTACTTGGTAAATCATGCTTATTCATCATCATCATCATCATCATCATCATCATCAT 156
QY 486 ProAspGlyLysLeuAlaLeuAlaAspPheGluAlaThrTrpLeuAlaMet 505
DB 157 CCAGATGGCTCAACATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 216
QY 506 AsnSerAspGluAspSerLysValAlaAlaThrTrpAspTrpGlyTrpGlnLeuGly 525
DB 217 AATACCCAGAGATGCTTCCCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 276
QY 526 MetAlaAspArgThrLeuValAspAsnThrTrpAsnThrHisLeuAlaLeu 545
DB 277 ATGGCTGATAGACACACTGTTGATACATATCATATCATATCATATCATATCATATCAT 336
QY 546 ValGlyLysAlaMetAlaSerProGluGluLysSerTrpGluLeuLysGluHisAsp 565
DB 337 GTTGTAAAGCAATGCTTCCCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 396
QY 566 ValAspTrpValLeuValIlePheGlyLeuLeuIleGlyPheGlyAspAspIleAsn 585
DB 397 GTTGATTAATGTTAGTTATTTTGGAGGGTTATTTGGGTTATTTCTGTTGATGATATAAC 456
QY 586 LysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluLeuLysGluArg 605
DB 457 AAATTCCTATGATGGTAAAGTATGCTGAGGATATCTGGCTGATGATGATGATGATGAT 516
QY 606 TyrPheTrpAlaGluGlyGlyTrpArgValAspAlaArgAlaSerGluThrMetArg 625
DB 517 GACTACTTACTGACCGGAGGAGATATAAAGTGGATTAAGATGATCATCATCTGGCAATGAAG 576
QY 626 AsnSerLeuLeuTrpLysMetSerTrpLysAspPheProGlnLeuPheAsnGlyGlyGln 645
DB 577 AATTCCTTGTATGATTAAGTATCTGATCATGATTTCAATGTTGTTGAGGTAGAT 636
QY 646 AlaThrAspArgValArgGlnMetIleThrProLeuAspValProProLeuAspTrp 665
DB 637 GGTGTTGATAGATTAGAAACCAAAATCCCAACCAATGAGATGACCAATGGAATGTT 696

QY 666 PheAspGluValPheThrSerGluAsnTrpMetValArgIleTrpGlnLeuLysAsp 685
DB 697 GTTCAAGAGAGCTTCATCATCAGAAATTTGATTTGAGAAATTTACAAATTAAGATTG 756
QY 686 AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer 702
DB 757 GATTAATGTTGATAGATTTTACATCAAGTACTGCTTTTGAAGAATCAATCATCCGGCAT 816
QY 703 ThrLysThrArgSerIleLysArgProGluLeuGlyLeuArg 717
DB 817 TCCAAAGAAACAGATCCATAAAGAGACCTAAATTCGAAGTAAGA 861

RESULT 3

US-09-270-767-11648
; Sequence 11648, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62617
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 11648
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-11648

Alignment Scores:
Pred. No.: 3.86e-98 Length: 1386
Score: 892.00 Matches: 172
Percent Similarity: 70.79% Conservative: 51
Best Local Similarity: 54.60% Mismatches: 82
Query Match: 23.72% Indels: 10
DB: 4 Gaps: 4

US-10-028-384-6 (1-718) x US-09-270-767-11648 (1-1386)

QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
DB 468 CTGTCAAGTGGCCATTTCTCATCTCGGACGCGTTTATCATTTGCCACACGCTTGTTC 527
QY 35 AlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAla 54
DB 528 TCTGTGTGGATTCGAAGCGTAATCCATGAGTTCGATCCGCTTCACTTCACTACCGCAC 587
QY 55 ThrLysTrpLeuValAsnAsnSerPheTrpLysPheLeuAsnTrpPheAspArgThr 74
DB 588 ACGGCGTTTCTGGCGGAGCAGGCGCTTTTACAGTTCCCACTGCTTCGATGACCGCGCC 647
QY 75 TrpTrpProLeuGlyArgValThrGlyGlyThrLeuTrpProGlyLeuMetThrThrSer 94
DB 648 TGGTATCCCTTGGGCCCATCATCGCGGACCATCTATCCCGGCGTGTGATGCTACCTCG 707
QY 95 AlaPheLeuTrpHisAlaLeuArgAsnTrp---LeuGlyLeuProIleAspIleArgAsn 113
DB 708 GCGGCCCTGTACCGCCGTGATG---TGGCTGCTCAATGTGACCATCGACATACGGAAC 761
QY 114 ValCysValLeuPheAlaProLeuPheSerGlyValIleAlaTrpAlaThrTrpGluPhe 133
DB 762 GTGTGCTCTCTCCGGGCCCTTCTCTCTCTCGCTGACCATCGCTGCTGCTACCGCCTC 821
QY 134 ThrLysGluLeuLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
DB 822 ACAAAGGAGATACACACACTGGAGCTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 881
QY 154 ProGlyTrpIleSerArgSerValAlaGlySerTrpAspAsnGluAlaIleAlaIleThr 173
DB 882 CCGCGGTATATCTCTCTGATCCGTGGGGGATCGTACGACAAATGAAGGCATCGCCATTTC 941

QY 532 LeuValAspSerThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetala 551
DB 917 CTAGTGATATATACGTGAACAAATAGTACATAGCGCTGGTTGGCAAGCAATGTCT 976
QY 552 SerProGluGluLysSerTrpGluIleLeuLysGluHisAspValAspTrpValLeuVal 571
DB 977 TCACCCGAGGAGAGCTTACGAATATGACATCTCTTGACGTGGACTAGCTTTGGTG 1036
QY 572 IlePheGlyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheIleTrpMetile 591
DB 1037 ATCTTTCCGCGTGTGATCGCTTATCTCGCATGATATCAACAAGTTCTCTGGATGGTC 1096
QY 592 ArgIleSerGluGlyIleTrpProGluGluIleLysGluArgTrpPheTrpAlaGlu 611
DB 1097 CGAATTCTCGAGGAGAGCATCCCAAGGACATTTAGGAAGCGATTACTTTACCGACGC 1156
QY 612 GlyCluTrpArgValAspAlaArgAlaSerGluThrMetArgAsnSerIleuLeuTylys 631
DB 1157 GGTGAATTCAGGTGATGCGGTGCGAAGTGTCTCGGCGCTCTCAACTGCTTATGTACAA 1216
QY 632 MetSerTrpLysAspPheProGluLeu-----PheAsnGlyGlyGlyLAlaThrAsp 648
DB 1217 TTAAGCTACTACAGATTCGGGAATTGAAGTTGGACTACAGAGTCCATCTCGATATGAT 1276
QY 649 ArgValArgGlnMetIleThrProLeuAspValProProLeuAspTrpPheAspGlu 668
DB 1277 CGCACAGTAAACCCGCTCATTTGGGAATAAGGACTTC---GATCTGACCTACCTCGAGGAG 1333
QY 669 ValPheThrSerGluAsnTrpMetValArgIleTrpGlnLeuLysLysAsp----- 685
DB 1334 GCCTACACCAACAGACACTGCTGTTGTCATCTATAGGTTAGGTTGAAGACCCGATGATTC 1393
QY 686 -----AspAlaGlnGlyArgTrpLeuArgAspValGlyGluLeuThrArg 700
DB 1394 AATAGACCATCACTGAAGACAGGAGAGACAGATTCCTCCAGCAAACTTCATTTCCGAGA 1453
QY 701 SerSerThrLysThrArgArgSerIleLysArg 711
DB 1454 AAGAACTCTAAGCTGCGCAAGGCGCTACATACGA 1486

RESULT 8

US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59. US2. REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.DM
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 122..505
US-09-513-999C-1965

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-----|
| Pred. No.: | 9,95e-63 | Length: | 507 |
| Score: | 595.00 | Matches: | 106 |
| Percent Similarity: | 81.44% | Conservative: | 30 |
| Best Local Similarity: | 63.47% | Mismatches: | 31 |
| Query Match: | 15.82% | Indels: | 0 |
| DB: | 4 | Gaps: | 0 |

US-10-028-384-6 (1-718) x US-09-513-999C-1965 (1-507)

QY 474 TyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIle 493
DB 2 TACTCTAGTCCCAAGTGTAGTCTCTGGCTCATACATCATGATGATGATGATGATGAT 61
QY 494 AspAspPheArgGluValaTrpTrpLeuArgMetAsnSerAspGluAspSerLysVal 513
DB 62 GATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGATGAACATGACCGAGTA 121
QY 514 AlaAlaTrpTrpAspTrpGlyTrpGlnIleGlyGlyMetAlaAspArgThrThrLeuVal 533
DB 122 ATGCTCTGGTGGGATTAATGCTATCAGATAGCTGGAATGGCTAATAGAACTAGCTGGTG 181
QY 534 AspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro 553
DB 182 GATAATACACCTGGAATACAGCCATAGCATGCTGGTGGAAAGCTATGCTTTCTTAAT 241
QY 554 GluGluLysSerTrpGluIleLeuLysGluHisAspValAspTrpValLeuValIlePhe 573
DB 242 GAAACAGCAGCCTATAAAATCATGAGGACTCTAGATGATGATGATGATGATGATGAT 301
QY 574 GlyGlyLeuIleGlyPheGlyGlyAspIleAsnLysPheLeuTrpMetIleArgIle 593
DB 302 GGAGGGGTATTTGGCTATTCTGTTGATGATATCAACAAATTTCTCTGGATGGTTAGGATA 361
QY 594 SerGluGlyIleTrpProGluGluIleLysGluArgTrpPheTrpThrAlaGluGlyGlu 613
DB 362 GCTGAAGAGAGAACATCCCAAGACATTCGGGAAGTCACTATTTTACCCACAGGAGAA 421
QY 614 TyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTrpLysMetSer 633
DB 422 TTCCCTGTAGCAAGACAGGATCCCTACTTTTGTGAATTCCTTATGATATAAATGTCA 481
QY 634 TyrLysAspPheProGlnLeu 640
DB 482 TACTACATTTGGAGAAATG 502

RESULT 9

US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12856
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-12856

Alignment Scores:

| | | | |
|------------------------|----------|---------------|-----|
| Pred. No.: | 7,43e-62 | Length: | 914 |
| Score: | 591.50 | Matches: | 114 |
| Percent Similarity: | 72.59% | Conservative: | 29 |
| Best Local Similarity: | 57.87% | Mismatches: | 43 |
| Query Match: | 15.73% | Indels: | 11 |
| DB: | 4 | Gaps: | 4 |

US-10-028-384-6 (1-718) x US-09-270-767-12856 (1-914)

QY 494 AspAspPheArgGluAlaTrpTrpLeuArgMetAsnSerAspGluAspSerLysVal 513
DB 5 GATGATTTCCGAGGCGCTACTACTGCTGATGATGAACATGACCTCCGAGGAGCGCTGCATA 64
QY 514 AlaAlaTrpTrpAspTrpGlyTrpGlnIleGlyGlyMetAlaAspArgThrThrLeuVal 533

Db 65 ATGTCCTGGTGGACTACGGCTACCGATACGATGCGCCATGCGCAATCGGACGATATTAGT 124
Qy 534 AspAsnAsnThrTriAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro 553
Db 125 GATAACAATACCTGGACACACATATATCGCGGTTGGCCAGGCGATGGCTCTTCG 184
Qy 554 GluGlySerThrGluLeuLeuLysGluHisAspValAspTyrValLeuValIlePhe 573
Db 185 GAGGAAGAGCCCTACGAGATAATGAGGGAACCTGGATGGGACTACGTTCTCGTATTTTC 244
Qy 574 GlyGlyLeuIleGlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleAglile 593
Db 245 GAGGGCTCACTGGCTACTCATCGGACGATATCAACAAGTTCCTGTGGATGGTGGCGATT 304
Qy 594 -----SerGluGlyIleTrpProGluGluIleLysGluAspTyrPheTyrThr 609
Db 305 GCGCGGACGCGATCGTGGT-----GCGCACATCCCGGAAAGGACTACTATGCG 355
Qy 610 AlaGluGlyGlyTyrArgValAlaAlaArgAlaSerGluThrMetArgAsnSerLeuLeu 629
Db 356 GCCAACGGAGAGTTCCGAGTGACAAAGGAGGCTCACCCACACACTGCTCAATTGTTTGATG 415
Qy 630 TyrLysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGly-----GlnAla 646
Db 416 TACAGATGTGCTACTATCGCTTGGGCAATGATACCGAAGTGGCAAGCCAGGCG 475
Qy 647 ThrAspArgValArgGlnMetIleThrProLeuAspValProProLeuAspTyrPhe 666
Db 476 TAGCATCGAGTTCTGTCGCCCGGAGATCGGCAACAGGACTTT---GAACCTGGATGCTCGT 532
Qy 667 AspGluValPheThrSerGluAsnTrpMetValArgIleTyrGlnLeuLys 683
Db 533 GAGGAGGCGTACACACGAGGACTGGCTGGTGGCGACTCTACAAAGTTAAG 583

RESULT 10

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Alignment Scores:
Pred. No.: 2,35e-51 Length: 1660
Score: 509.00 Matches: 101
Percent Similarity: 68.00% Conservative: 35
Best Local Similarity: 50.50% Mismatches: 52
Query Match: 13.53% Indels: 12
DB: 4 Gaps: 3

US-10-028-384-6 (1-718) x US-09-270-767-158 (1-1660)

Qy 523 IleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTriAsnAsnThrHis 542
Db 796 GTTGGCGGAATGCGCAACAGACGCGCTAGTGGGATAATAATACGTGGAAACAATAGTCAC 737
Qy 543 IleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLys 562
Db 736 ATAGCGCTGGTGGCAAGCAATGCTCTCAACCGAGGAGAGTCTCTACGAAATATGACA 677
Qy 563 GluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAsp 582
Db 676 TCTCTTGACGTGACTACGTTTGGTGATCTTTGGCGGTGTGATCGCTATTCCGCGGAT 617

Qy 583 AspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIle 602
Db 616 GATATCAACAAGTTCCTGTGGATCGGATTCGGAATTCGCGAGGAGAGCATCCCAAGGACATT 557
Qy 603 LysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAlaSerGlu 622
Db 556 AAGGAAGCGATTACTTTCCGACCGCGGTGAATTCAGGTAGATGCCGAAGTCTCGG 497
Qy 623 ThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu----- 640
Db 496 GCCCTGTCTCACTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATGAAGTTG 437
Qy 641 ---PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAsp 659
Db 436 GACTACAGAGCCCATCTGGATATGATCGACACAGTAACCGCTCATTTGGGAATAAGGAC 377
Qy 660 ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIle 679
Db 376 TTT---GATCTGACCTACCTCGGAGGAGGCTACACACAGACACTGGCTTCTTCGCATC 320
Qy 680 TyrGlnLeuLysLysAsp-----AspAlaGlnGlyArgThr 691
Db 319 TATAGTGTGAGAGAGCGCGCATGAGTTCATAGACCATCACTGAAGACCAAGGAGAGACG 260
Qy 692 LeuArgAspValGlyGluLeuThrArgSerSerThrIleThrArgArgSerIleLysArg 711
Db 259 ATTCTCCAGCAAACTTCATCTCGAAGAAAGAACTCGAAGCGTCCGAAGGCTACATACGA 200

RESULT 11

US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 2,35e-51 Length: 1660
Score: 509.00 Matches: 101
Percent Similarity: 68.00% Conservative: 35
Best Local Similarity: 50.50% Mismatches: 52
Query Match: 13.53% Indels: 12
DB: 4 Gaps: 3

US-10-028-384-6 (1-718) x US-09-270-767-15440 (1-1660)

Qy 523 IleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTriAsnAsnThrHis 542
Db 796 GTTGGCGGAATGCGCAACAGACGCGCTAGTGGGATAATAATACGTGGGAACAATAGTCAC 737
Qy 543 IleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLys 562
Db 736 ATAGCGCTGGTGGCAAGCAATGCTCTCAACCGAGGAGAGTCTCTACGAAATATGACA 677
Qy 563 GluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAsp 582
Db 676 TCTCTTGACGTGACTACGTTTGGTGATCTTTGGCGGTGTGATCGCTATTCCGCGGAT 617
Qy 583 AspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIle 602
Db 616 GATATCAACAAGTTCCTGTGGATCGGATTCGGAATTCGCGAGGAGAGCATCCCAAGGACATT 557

QY 603 LysGluArgTyrPheTyrThAlaGluGlyClnTyrArgValAspAlaArgAlaSerGlu 622
DB 556 AAGGAAAGCGATTAATCTTACCGACCGCGGTGAATTCAGGTAGATCCGAAGTGTCTCG 497
QY 623 ThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu----- 640
DB 496 GCCCTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTG 437
QY 641 ---PheAsnGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAsp 659
DB 436 GACTACAGAGGCCCATCTCGATATGATCGCACAGCTAACCGCTCATTCGGGAATTAAGGAC 377
QY 660 ValProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgile 679
DB 376 TTT---GATCTGACCTACCTCGAGGAGGCTTACACACACAGCACTGGCTTGTTCGATC 320
QY 680 TyrGlnLeuLysLysAsp-----AspAlaGlnGlyArgThr 591
DB 319 TATAGTGTGAAGACCCGATGATCAATAGACCATCTAGTGAAGACCAAGGAGAGACG 260
QY 692 LeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLysArg 711
DB 259 ATTCTCCAGCAAACTTCTATCTCGAAGAAAGAACTCGAAGGCTCGCAAGGCTTACATAGA 200

RESULT 12
US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 27262
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-27262

Alignment Scores:
Pred. No.: 1,786-46 Length: 503
Score: 461.50 Matches: 93
Percent Similarity: 70.11% Conservative: 29
Best Local Similarity: 53.45% Mismatches: 45
Query Match: 12.27% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-6 (1-718) x US-09-270-767-27262 (1-503)
QY 155 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrIleu 174
DB 2 GGGTATATCTCGATCGTGGCGGATCTGACAAATGAAGGATCCCATTTCTGCG 61
QY 175 LeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMethisala 194
DB 62 ATGCTCTTCACTACTATTGTCGATCAAGCGGTAAAGCGGACGATCTTTGGTCG 121
QY 195 ThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 214
DB 122 GCTATGCGGATGGCGATCTTCTATGCTCTCTCGGGGTCGATGCTCTCTG 181
QY 215 ThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSerLys 234
DB 182 ATTAACCTAATCCCGCTGCAGCTCTCGGCTGATGATCACCGACGTTCTCGCACAC 241
QY 235 LeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIlePro 254
DB 242 ATCTACATAGATACAGACGCTATCTGCTGGACCACTTCTGTCGATGACATCTCG 301
QY 255 PheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGly 274

DB 302 TTGTGGGATTCACACCATCCAGAGCTCCGAACACATGCTGGCATCGGAACCTTTGGC 361
QY 275 LeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysPhe 294
DB 362 CTGTGCCAGATTACGCTTTGTCGACTATCTGCGCTCGCGCATCCCAAGGATCACTTC 421
QY 295 LysValIleMetMetValSerIlePheLeuIleLeuValLeuGlyValValGlyLeuSer 314
DB 422 GATCTG-----CTCTTCAAGACGTTGGTTTCCAGTGT-----TTGACT 460
QY 315 AlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArg 328
DB 461 GTGGTGTTCGTGCTGGTACCTGCTCAGCGCTTACCGGAAA 502

RESULT 13
US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll, III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 332
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)---(616)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

Alignment Scores:
Pred. No.: 6,396-45 Length: 616
Score: 450.00 Matches: 90
Percent Similarity: 65.52% Conservative: 24
Best Local Similarity: 51.72% Mismatches: 54
Query Match: 11.96% Indels: 6
DB: 3 Gaps: 3

US-10-028-384-6 (1-718) x US-09-328-111-332 (1-616)
QY 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIle 172
DB 5 GTACAGAGCTACATATCTCGGTAGTGGATCTTTGATATGAAGCATTTGCTATT 64
QY 173 ThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMet 192
DB 65 TTTCACCTTCAGTTCACATCTATTTATGGTAAATCTGTAAAACTGGGTCAGTTT 124
QY 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVal 212
DB 125 TGGACAATGTGCTGCTTATCTATATGCTCTCTCTGCTTGGGGTGGTTATGTA 184

| | | | |
|----|-----|---|-----|
| Qy | 213 | PhaileThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSer | 232 |
| Db | 185 | TTTTATCATCAATCTTATTCACCTGATGATTTGTTGTACTGATGCAGATACAGC | 244 |
| Qy | 233 | SerLysLeuTyrSerAlaTyrThrThrTyrValalIleGlyThrValAlaSerMetGln | 252 |
| Db | 245 | AAAAGAGCTTACATAGCATATAGCACTTTCACATTGTGGGTTTAATATTATCAATGCAG | 304 |
| Qy | 253 | IleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaIleLeuGlyVal | 272 |
| Db | 305 | ATACCTTTTGGGATTCCAGCAATCAGAACCAAGTGAACACATGCGCAGCTTCAGGTGC | 364 |
| Qy | 273 | PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla | 292 |
| Db | 365 | TTTGCAATTGCTCGAAGCTTAANCCTTCTTGCGATATCTGAGAACCGATTACCACCAACCAAG | 424 |
| Qy | 293 | LysPheLeuValIleValMetMetValSerLeuPheLeuIleLeuValLeuGlyValValGly | 312 |
| Db | 425 | AGTTCAGACCCCTTCNTTTTGGGGATACATAC-----TTTCAGNGCTGGTCTCTTANGGC | 478 |
| Qy | 313 | LeuSerAlaLeuThrTyrMetGlyLeuIleAlaPro---Trp | 325 |
| Db | 479 | ATTATTGNTATTCNGTAC-----ATTCCCTCGGATGG | 511 |

```

RESULT 14
US-09-385-982-213
; Sequence 213, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-213

```

| | | |
|-------------------------|----------|---------------------------|
| Alignment Scores: | | |
| Pred. No.: | 4.18e-40 | Length: 487 |
| Score: | 409.00 | Matches: 78 |
| Percent Similarity: | 70.19% | Conservative: 35 |
| Best Local Similarity: | 48.45% | Mismatches: 40 |
| Query Match: | 10.97% | Indels: 8 |
| DB: | 3 | Gaps: 3 |
| US-10-028-384-6 (1-718) | x | US-09-385-982-213 (1-487) |

[illegible]

| | | | |
|----|-----|---|-----|
| Db | 181 | GTCACTCTCTGGTAGCGTTTGTCTTCTCAACCGTGGAGCGCTCTC-----CTC | 228 |
| Qy | 317 | ThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsn | 336 |
| Db | 229 | ATGCTGACAGGAAAAATATCTCCCTGGACGGGGGTTTCTACTCACTGCTGGATCCCTCT | 288 |
| Qy | 337 | TyrAlaValIleHisIleSerPheIleIleAlaSerValSerGluHisGlnProValSerTrp | 356 |
| Db | 289 | TATGCTTAGAACAACATCCCATCATTCCTTCTGTCTGAGCATCAGCCCAACACTGG | 348 |
| Qy | 357 | ProAlaPhePheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeu | 376 |
| Db | 349 | TCCTCATACTATTATGACCTGCAGCTCCTCGCTCTTCATGTTCCAGTTGSCCCTCTATTAC | 408 |
| Qy | 377 | LeuPheLeuAspLeuIleAspGluHisValPheValIleAlaTyrSerValLeuCysSer | 396 |
| Db | 409 | TGCTTTAGCAACCTGTCTGATGCCCGGATTTTATCATCATGTATGCTGTGACCGCATG | 468 |
| Qy | 397 | Tyr | 397 |
| Db | 469 | TAC | 471 |

```

RESULT 15
US-09-702-705-1655
; Sequence 1655, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaityana S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Vannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version
; SEQ ID NO 1655
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-1655

```

| | | |
|-------------------------|---------|----------------------------|
| Alignment Scores: | | |
| Pred. ID: | 1.6e-38 | Length: |
| Score: | 396.00 | Matches: |
| Percent Similarity: | 83.50% | Conservative: |
| Best Local Similarity: | 69.90% | Mismatches: |
| Query Match: | 10.53% | Indels: |
| DB: | 4 | Gaps: |
| US-10-028-384-6 (1-718) | x | US-09-702-705-1655 (1-487) |

| | | | |
|----|-----|---|-----|
| QY | 473 | AlatyrSerSerProCserValValLeuProSerGlnThrProAspGlyLysLeuAlaLeu | 492 |
| | ::: | ::::: | ::: |
| Db | 74 | TCCCTACTTCTTCGGTCCATTGTACTATCTGCCCGTGTTGGGAGTCGCAGTAGGATCATTA | 133 |
| | | ::::: | |
| QY | 493 | IleAspAspPheArgGluAlaTyrrTyrrTrrLeuArgMetAsnSerAspGluAspSerLys | 512 |
| | | ::::: | |
| Db | 134 | TTTGATGACTTCGAGAAGCATATTATTGGCTCGTCATATACTCCAGAGGATCGCGAG | 193 |
| | | ::::: | |
| QY | 513 | ValIalaLaTrpTrrAspTyrrGlyrGlyrGlnIleGlyrGlyMetAlaAspArgThrThrrLeu | 532 |
| | ::: | ::::: | |
| Db | 194 | GTCATGTCTCGTGGGATTATGGCTATCAGATTACAGCTATGGCAACCGAACAAATTITTA | 253 |
| | | ::::: | |
| QY | 533 | ValAspAenAenThrTrrAenAenThrHisIleAlaIleValGlyrLysAlaMetAlaSer | 552 |
| | | ::::: | |

us-10-028-384-6.rni

Thu Dec 16 16:25:10 2004

| | | | |
|----|-----|--|-----|
| Db | 254 | GTGGCAATAACACATGGGAATAATACCCATATTTCTCGATAGGCGAGCAATGGCGTCC | 313 |
| Qy | 553 | ProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValile | 572 |
| Db | 314 | ACAGAGGAAAGCCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTCGTGTCATT | 373 |
| Qy | 573 | PheGlyGly | 575 |
| Db | 374 | TTTGAGGA | 382 |

Search completed: December 15, 2004, 11:59:00
Job time : 170.657 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: December 15, 2004, 06:02:06 : Search time 903.972 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761

Sequence: 1 MGSRRSCVLSVFQTKLVI.....TRSTKTRRSIKRPELGLRV 718

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pz2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_epool/US10028384/runat_14122004_131523_8140/app_query.fasta_1.3740
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rpnb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10028384 -CGN 1 1 2577 @runat_14122004_131523_8140
-NCPUS=6 -ICPU=3 -NO YMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLCK=100
-JONGLOG -DELT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 3761 | 100.0 | 2733 | 15 | US-10-028-384-5 |
| 2 | 3751 | 99.7 | 2157 | 9 | US-09-801-368-387 |
| 3 | 3751 | 99.7 | 2157 | 18 | US-10-793-639-318 |
| 4 | 2738 | 72.8 | 2256 | 15 | US-10-032-585-6323 |
| 5 | 2438.5 | 64.5 | 2232 | 15 | US-10-128-714-7139 |
| 6 | 2301.5 | 61.2 | 2466 | 16 | US-10-320-797-2305 |
| 7 | 2284.5 | 60.2 | 2603 | 15 | US-10-128-714-1139 |
| 8 | 2264.5 | 60.2 | 3669 | 15 | US-10-128-714-1139 |
| 9 | 2264.5 | 60.2 | 4603 | 15 | US-10-128-714-5139 |
| 10 | 2090 | 55.6 | 1848 | 15 | US-10-128-714-2139 |
| 11 | 2042 | 54.3 | 1969 | 15 | US-10-128-714-1139 |
| 12 | 2032.5 | 54.0 | 2882 | 16 | US-10-320-797-1305 |
| 13 | 2032.5 | 54.0 | 4738 | 16 | US-10-320-797-305 |
| 14 | 2037.5 | 53.9 | 2417 | 15 | US-10-028-384-7 |
| 15 | 2013.5 | 53.5 | 2710 | 15 | US-10-028-384-3 |
| 16 | 2012.5 | 53.5 | 2481 | 15 | US-10-028-384-1 |
| 17 | 1946 | 51.7 | 3046 | 18 | US-10-417-375-95 |
| 18 | 1945 | 51.7 | 3093 | 18 | US-10-417-375-92 |
| 19 | 1945 | 51.7 | 3094 | 15 | US-10-028-384-9 |
| 20 | 1934 | 51.4 | 5404 | 18 | US-10-417-375-99 |
| 21 | 1934 | 51.4 | 5827 | 18 | US-10-417-375-97 |
| 22 | 1920 | 51.1 | 2472 | 15 | US-10-171-581-112 |
| 23 | 1920 | 51.1 | 2472 | 15 | US-10-028-384-11 |
| 24 | 1920 | 51.1 | 2472 | 15 | US-10-172-118-742 |
| 25 | 1920 | 51.1 | 2472 | 16 | US-10-342-887-742 |
| 26 | 1754.5 | 46.6 | 2681 | 16 | US-10-425-114-1408 |
| 27 | 1754 | 46.6 | 2839 | 18 | US-10-425-115-150745 |
| 28 | 1739 | 46.2 | 2779 | 17 | US-10-437-963-99904 |
| 29 | 1684 | 44.8 | 3197 | 16 | US-10-424-593-111541 |
| 30 | 1680.5 | 44.7 | 3141 | 18 | US-10-425-115-130787 |
| 31 | 1402 | 37.3 | 2244 | 17 | US-10-437-963-48342 |
| 32 | 1156 | 30.7 | 1543 | 10 | US-09-974-879-133 |
| 33 | 1156 | 30.7 | 1543 | 10 | US-09-305-736-133 |
| 34 | 1156 | 30.7 | 1543 | 11 | US-09-818-683-133 |
| 35 | 1156 | 30.7 | 1543 | 11 | US-09-818-683-133 |
| 36 | 1156 | 30.7 | 1543 | 16 | US-10-621-401-133 |
| 37 | 1154 | 30.7 | 1209 | 15 | US-10-106-698-330 |
| 38 | 1154 | 30.7 | 1209 | 16 | US-10-264-237-412 |
| 39 | 1000 | 26.6 | 1728 | 16 | US-10-424-593-122476 |
| 40 | 885 | 23.5 | 1094 | 17 | US-10-437-963-99902 |
| 41 | 796 | 21.2 | 1828 | 10 | US-09-945-527-62 |
| 42 | 739 | 19.6 | 1114 | 16 | US-10-296-115-629 |
| 43 | 719.5 | 19.1 | 2660 | 16 | US-10-264-049-630 |
| 44 | 661 | 17.5 | 887 | 16 | US-10-424-593-77697 |
| 45 | 561.5 | 14.9 | 904 | 17 | US-10-767-701-9050 |

ALIGNMENTS

RESULT 1
US-10-028-384-5
; Sequence 5, Appli US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D28952
; DATABASE ENTRY DATE: 1999-02-07

RELEVANT RESIDUES: (1)...(2733)

US-10-028-384-5

Alignment Scores:

Pred. No.: 0
 Score: 3761.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 Indels: 0
 Gaps: 15

US-10-028-384-6 (1-718) x US-10-028-384-5 (1-2733)

Qy 1 MetGlySerAspArgSerCysValLeuSerValPheGlnThrIleLeuLysLeuValIle 20
 Db 219 ATGGATCCGACCGGCTGGTGTGTTGCTGTGTTTCAGACCATCTCTCAAGCTCGTCATC 278
 Qy 21 PheValAlaIlePheGlyAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
 Db 279 TTCGTGGCGATTTTGGGGGTGCGCATATCATCACGTTTGTTCGACGATCAAAATTTGAG 338
 Qy 41 SerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrLysTyrLeuValAsn 60
 Db 339 TCATATTCATGATTCGACCCCTGGTTCATATATAGGGCTACCAATATCTCGTCAAC 398
 Qy 61 AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTyrTrpProLeuGlyArg 80
 Db 399 AATTCGTTTACAGATTTTGAACCTGGTTGACGACCGTACCTGGTACCCCTCGGAAGG 458
 Qy 81 ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100
 Db 459 GTTACTCGAGGACTTATATCTCTGTTGATGACGACTAGTGGCTTCATCTGCGACGCC 518
 Qy 101 LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
 Db 519 CTGGCAACTCGTGGCTTGGCCATTGACATGACAGAAAGTGTGTGCTATTTCGGCCA 578
 Qy 121 LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluIleLysAspAla 140
 Db 579 CTATTTCTGGGTCACCGCTGGCGACTTACGAAATTTACGAAAGAGATTAAGATGCC 638
 Qy 141 SerIleGlyLeuLeuAlaIleGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
 Db 639 AGCCCTGGGTTTGGCTGTGTTTATAGCCATTGTCCCGGTTATATCTAGATCA 698
 Qy 161 ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
 Db 699 GTGCGGGTCTTACGATATAGGCCATTGCCATTACACTATTATATGTCACCTTTCAATG 758
 Qy 181 PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeuphe 200
 Db 759 TTTTGGATTAAGGCCCAAGATGTGCTCTATCATGACGCAACGTGTGCGACTTTATTC 818
 Qy 201 TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
 Db 819 TACTTCTACATGTGTGCTGGCTGGGTGGATACGTTTTCATCAACCACTTGATCCCATC 878
 Qy 221 HisValPheLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyrThr 240
 Db 879 CATGCTCTTTTGTGATTTTGAAGGGCAGATATTCGTCCAAACTGTATTCGCTACACC 938
 Qy 241 ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
 Db 939 ACTTGTGTACGCTATTGGAATCTTGTGATCCATGCGAGATCCCATTTTGTGGTTTCTACCT 998
 Qy 261 IleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
 Db 999 ATCAGGCTTACGACCAATGGCCGATTTGGGTGTTTTCGGTTTGAATCAGATGTGCGCC 1058
 Qy 281 PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
 Db 1059 TTCGGTGACTCGTGAAGGGCCAAATCAGCACAGCTTAAGTTTAAAGTTCATCATGATGTT 1118

Qy 301 SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
 Db 1119 TCTCTGTTTGTGATCTTGGTCTCTGGTGTGGTGGACTTTCTGCCTTGACCTATATATGGGG 1178
 Qy 321 LeuIleAlaProTyrThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
 Db 1179 TTGATTCGCCCTTGGACTGGTAGATTATTTCGTTATGGGATACCAACTAGCCAAAGATC 1238
 Qy 341 HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProIlePhePhe 360
 Db 1239 CACATTCCTATCATTTGCCCTCCGTTTCGGAACATCAACCCGTTTCGTGGCCGCTTTCTTC 1298
 Qy 361 PheAspThrHisPheLeuIleTrpPheProAlaGlyValPheLeuLeuPheLeuAsp 380
 Db 1299 TTTGATACCCACTTTTGTATCTGGCTATTCGCCGCGGTGTAATCTCTACTATTCTCTCGAC 1358
 Qy 381 LeuLysAspGluHisValPheValIleAlaTyrSerValLeuLysSerTyrPheAlaGly 400
 Db 1359 TTGAAAGACGACGACGATTTTGTTCATCGCTTACTCCGTTCTGTGTGCTACTTTGCGCGT 1418
 Qy 401 ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
 Db 1419 GTTATGTTAGATTGATGTTGACTTTTGACACCACTCATCTGTGTGCGCGCGCTCGCA 1478
 Qy 421 LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
 Db 1479 TTGTCCAAAGATTTTGACATCTACTCTGGATTTCAGACAAAGTACCGCAAAATACGCCATC 1538
 Qy 441 LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
 Db 1539 AAACCTGCGGCACCTACTGCGCAAAATTTGATGTTTCCGATCATCTATTTTATTGTAT 1598
 Qy 461 LeuPheValPheHisSerThrTyrValThrArgThrAlaTyrSerSerProSerValVal 480
 Db 1599 CTTTTGCTTCCACTTCTACTTGGTAAACAGAACTGATCATCTCTCTCTCTCTCTCTCT 1658
 Qy 481 LeuProSerGlnTrpProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr 500
 Db 1659 TTGCCATCACAACCCAGATGTTAAATTTGGCGTTGATCGACGACTTCAGGGAAGCGTAC 1718
 Qy 501 TyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGly 520
 Db 1719 TATTGGTTAAGAATGAATCTGTATGAGCAGATGAAGTTTTCAGCGTGGTGGGATACGGT 1778
 Qy 521 TyrGlnIleGlyLysMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 540
 Db 1779 TACCAAAATTTGGTCATGGCAGACAGAACCACTTTAGTCGATTAACAACACGTTGGAACAAT 1838
 Qy 541 ThrHisIleAlaIleValGlyValAlaMetAlaSerProGluGlySerTyrGluIle 560
 Db 1839 ACTCATCTGCCCATCTGTGTGTAAGCCATGGCTTCCCTTGAAGAGAAATCTTACGAAT 1898
 Qy 561 LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly 580
 Db 1899 CTAAAGAGCATGATGTCGAATATGTCCTCATCTTTGGTGTCTATAATTTGGTTTGT 1958
 Qy 581 GlyAspAspIleAsnLysPheLeuTrpValIleArgIleSerGluGlyIleTrpProGlu 600
 Db 1959 GGTGATGATCATCAAAATTTCTGTGATGATCAGAAATAGCAGGGAATCTGCCAGAA 2018
 Qy 601 GluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAspAlaArgAla 620
 Db 2019 GAGATAAAGAGCGTTATTTCTATACCGCAGAGGAGATACAGATGATGATGCAAGGCT 2078
 Qy 621 SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu 640
 Db 2079 TCTGAGACCATGAGGAATCTGCTACTTTACAAATGTCCTACAAAGATTTTCCCAATTA 2138
 Qy 641 PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal 660
 Db 2139 TTTCAATGTTGGCCAAAGCCACTGACAGAGTGGTCAACAAATGATCATCACCATTAGACGTC 2198
 Qy 661 ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr 680

```

Db      2199  CCACATTAGACTACTTCGAGAAAGTTTCTACCTTCGANAACCTGGATGGTAGATATAT 2258
Qy      681  GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg 700
Db      2259  CAATTGAAGAAGGATGATGTCACCAAGGTAGAACTTTGAGGGACGTTGGTGAGTTAACACAG 2318
Qy      701  SerSerThrLysThrArgArgSerLysLysArgProGluLeuGlyLeuArgVal 718
Db      2319  TCTTCTACGAACACAGAAAGGTCCATAAAGACCTGAATAGGCTTGAGAGTC 2372

RESULT 2
US-09-801-368-387
; Sequence 387, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Calli, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofia
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 387
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-387

Alignment Scores:
Pred. No.:      0      Length:      2157
Score:          3751.00  Matches:      717
Percent Similarity: 99.86%  Conservatives: 0
Best Local Similarity: 99.86%  Mismatches: 1
Query Match:      99.73%  Indels:      0
DB:               9      Gaps:      0

US-10-028-384-6 (1-718) x US-09-801-368-387 (1-2157)

Qy      1  MetGlySerAspArgSerCysValLeuSerValPheGlnThrLysLeuValIle 20
Db      1  ATGGGATCCGACCGGTGGTGTGTTTGTCTGTGTTTCAGACCAATCCCAAGCTCGTCATC 60
Qy      21  PheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPheAlaValIleLysPheGlu 40
Db      61  TTCGTGGCGAATTTTGGGGGTGCCATATCATCACGTTTGTTCAGTCATCAAAATTTGAG 120
Qy      41  SerLleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrLysTyrLeuValAsn 60
Db      121  TCTATTATCCATGAATTCGACCCCTGGTTCAATTATATAGGGCTACCAAAATATCTCGCAAC 180
Qy      61  AsnSerPheTyrLysPheLeuAsnTrpPheAspArgThrTrpTyrProLeuGlyArg 80
Db      181  AATTCGTTTACAAGTTTGAACCTGGTTTGCACGACCGTACCTGGTACCCCTCGGAGG 240
Qy      81  ValThrGlyGlyThrLeuTyrProGlyLeuMetThrThrSerAlaPheIleTrpHisAla 100

```

```

Db      241  GTTACTCGAGGAGCTTTATATCTCTGTTTCATGACGACTAGTCGCTTCACTTGGCAGCGC 300
Qy      101  LeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysValLeuPheAlaPro 120
Db      301  CTGGCAACTGGTGGGCTTGCCTCAITGACATCAGAAACGTTTGTGTGCTATTTGGCGCA 360
Qy      121  LeuPheSerGlyValThrAlaTrpAlaThrTyrGluPheThrLysGluLysAspAla 140
Db      361  CTATTTTCGGGTCACCGCTGGGCGACTTACGANTTTACGAAAGAGATTAAAGATGCC 420
Qy      141  SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyrIleSerArgSer 160
Db      421  AGCGTGGGCTTTTGGCTGTGTGTTTATAGCCATTGTCGCCGTTATATATATCATGATCA 480
Qy      161  ValAlaGlySerTyrAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPheMet 180
Db      481  GTGGGGGCTCTACGATAATAGGCCATTTGCCATTACACTATTAATGTCACCTTTTCATG 540
Qy      181  PheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaLeuPhe 200
Db      541  TTTTGGATTAAAGGCCCAAGAGACTGGCTCTATCATGCGCAACGTCGTGCAGCTTTTATTC 600
Qy      201  TyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuIleProLeu 220
Db      601  TACTTCTACATGGTGGCTGGGGTGGATAGCTGTTTCATCACCAACTTTGATCCCACTC 660
Qy      221  HisValPheLeuLeuIleLeuMetGlyArgTyrSerSerLysLysLeuTyrSerAlaTyr 240
Db      661  CATGCTCTTTTGGCTGATTTTGTATGGGCGAGATATTCGTCCAAAGCTGATTTCTGCACAC 720
Qy      241  ThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuPro 260
Db      721  ACTTGGTACGCTATGGAACTGTTGCATCCATGCGATCCCATTTGCGTTTCCATCCT 780
Qy      261  IleArgSerAsnAspHisMetAlaLeuGlyValPheGlyLeuIleGlnIleValAla 280
Db      781  ATCAGGCTTAACGACCACTATGGCGCATTTGGGTGGTGTGTTTTCGTTTGTATCAGATT 840
Qy      281  PheGlyAspPheValLysGlyGlnIleSerThrAlaLysPheLysValIleMetMetVal 300
Db      841  TTCGGTGACTTCGTGAAGGSCCAATCAGCAGCAGCTAAGTTTAAAGTCATCATGAGTT 900
Qy      301  SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly 320
Db      901  TCTCTGTTTGTGATCTTGGTCTTGGTGGTGGTGGTCTTCTGCTGTGACCTATATATGG 960
Qy      321  LeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAspThrAsnTyrAlaLysIle 340
Db      961  TTGATTGGCCCTTGGACTGGTAGATTTTATTCGTTATGGGATACCAACTACGCAAGATC 1020
Qy      341  HisIleProIleIleAlaSerValSerGluHisGlnProValSerTrpProAlaPhePhe 360
Db      1021  CACATTCTATCATTTGCTCGCTTTCGAAACATCAACCCCGTTTTCGTGGCCCGCTTCTTC 1080
Qy      361  PheAspThrHisPheLeuIleTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAsp 380
Db      1081  TTTGATACCCACTTTTGTACTGCTATTTCCCGCGGIGTATTCCTACTATTTCTCTGAC 1140
Qy      381  LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly 400
Db      1141  TTAAAGACGAGCAGCAGCTTTTGTCTATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 1200
Qy      401  ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla 420
Db      1201  GTTATGGTTAGATTGATTGACTTTTGACACCAAGTCTATCTGTGTGTCCCGCGCGCTGCA 1260
Qy      421  LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle 440
Db      1261  TTGTCCAAGATATTGACATCTACCTGGAATTCAGACAAAGTACCGCAAAATACGCCATC 1320
Qy      441  LysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr 460
Db      1321  AAACCTGGGCACACTACTGGCCAAATGATTGTTTCCGGATCATTCATCTTTATTTGAT 1380

```

| | | | | |
|----|--|------|---|------|
| QY | | 461 | LeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSerValVal | 480 |
| DB | | 1381 | CTTTTCGTCTCCATCTACTTGCGTAACAAGAACTGCATACTCTCTCCTCTCTGTTGT | 1440 |
| QY | | 481 | LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheAtgGlualaTyr | 500 |
| DB | | 1441 | TTCGCACTCACAAACCCAGATGGTAATTGGCGTTCATCGACGCTTCAGGAAGCGTAC | 1500 |
| QY | | 501 | TyrTrpLeuArgMetAenSerAspGluAspSerLysValAlaAlaTrpTrpAspTyrGly | 520 |
| DB | | 1501 | TATTGGTTAAGAATGAATCTGATGAGGACAGTAGGTTCAGCGCTGGTGGCATACGGT | 1560 |
| QY | | 521 | TyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsn | 540 |
| DB | | 1561 | TACCANAATTGGTGCACTGGCAGACAGAACCACTTTAGTCGATAAACAACCTGGAAACAAT | 1620 |
| QY | | 541 | ThrHisIleAlaIleValGlyLysAlaMetalaserProGluGluLysSerTyrGluIle | 560 |
| DB | | 1621 | ACTCACATCGCCATCGTTGGTAAGCCATGGCTTCCCCTGAAGAGAAAATCTTACGAATT | 1680 |
| QY | | 561 | LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleSGLyPheGly | 580 |
| DB | | 1681 | CTAAAGAGCATGATGCGAATTATGCTTGGTCATCTTTGGTGGTCTAAATGGGTTTTGGT | 1740 |
| QY | | 581 | GlyAspAspIleAsnLysPheLeuTyrMetIleAerGilleserGluGlyLyletrProGlu | 600 |
| DB | | 1741 | GGTGNATGACATCAACAATTTCTTTGTGGATGATCAGNAATTAGCGAGGGAAATCTGCCCAAGAA | 1800 |
| QY | | 601 | GluLeyLysGluArgTyrPheTyrThralaGluGlyGlyTyrArgValAspAlaArgAla | 620 |
| DB | | 1801 | CAGATAAAAGAGCGTGATTCTATACCGCAGAGGGAGAAATACAGAGTAGATAGTCAAGGGCT | 1860 |
| QY | | 621 | SerGluThrMetArgAenSerLeuLeuTyrLysMetSerTyrLysaspPheProGlnLeu | 640 |
| DB | | 1861 | TCTGAGACCATGAGGAACCTCGTACTTTTACAGATGCTCTCAAGAATTTCCCCAATTA | 1920 |
| QY | | 641 | PheAsnGlyGlyGlnAlaThrAspArgValArgGlnMetIleThrProLeuAspVal | 660 |
| DB | | 1921 | TTCAATGTGTGCCCAAGCCA CTGACAGAGTGGCTCAACAAATGATCACACCATTAGACGTC | 1980 |
| QY | | 661 | ProProLeuAspTyrPheAspGluValPheThrSerGluAenTyrMetValArgIleTyr | 680 |
| DB | | 1981 | CCACCATTAGACTACTTTCGACGAAGTTTTTACTTCGAAACCTGGATCGGTTAGAAATATAT | 2040 |
| QY | | 681 | GlnLeuLysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg | 700 |
| DB | | 2041 | CAATTGAAGAGGATGATGCCAAGTAGAATTTTGAGGGACGTTGGTGGATTAAACCAGG | 2100 |
| QY | | 701 | SerSerThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal | 718 |
| DB | | 2101 | TCCTTCTAGAAAAACAGAGAGTCCATAAGAGACCTGAATAGGTTGAGGCTCAGAGTC | 2154 |

RESULT 3

```

1  RESULT 3
2  US-10-793-639-318
3  / Sequence 318, Application US/10793639
4  / Publication No. US20040199940A1
5  / GENERAL INFORMATION:
6  / APPLICANT: Karunamandaa, Balasulsojini
7  / APPLICANT: Yu, Jaehyuk
8  / APPLICANT: Kishore, Ganesh M.
9  / TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
10 / WITH STEROL SYNTHESIS AND METABOLISM
11 / FILE REFERENCE: 16516.075
12 / CURRENT APPLICATION NUMBER: US/10/793,639
13 / CURRENT FILING DATE: 2004-03-05
14 / PRIOR APPLICATION NUMBER: US/03/614,221A
15 / PRIOR FILING DATE: 2000-07-11
16 / PRIOR APPLICATION NUMBER: US 60/142,981
17 / PRIOR FILING DATE: 1999-07-12
18 / NUMBER OF SEQ ID NOS: 626
19 / SEQ ID NO 318
20 / LENGTH: 2157

```

| | | | |
|----|------|--|------|
| QY | 301 | SerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyrMetGly | 320 |
| DB | 901 | TTCTCTGTTTGTGATCTTGGTCCCTGTGGTGTGGACTTCTCTGCTCTGACCTATATGGG | 960 |
| QY | 321 | LeuIleAlaProTTPThrGlyArgPheTyrSerLeuTTPAspThrAsnTyrAlaLysIle | 340 |
| DB | 961 | TTGATTGGCCCTTGGACTGGTAGATTTTATTGGTTATGGGATACCACTACGCAAGATC | 1020 |
| QY | 341 | HisIleProIleIleAlaSerValSerGluHisGlnProValSerTTPProAlaPhePhe | 360 |
| DB | 1021 | CACATTCCTATCATTGCTCCGTTTCGAAACATCAACCCGTTTCGTGGCCGCTTCTTC | 1080 |
| QY | 361 | PheAspThrHisPheLeuIleTTPLeuPheProAlaGlyValPheLeuLeuPheLeuAsp | 380 |
| DB | 1081 | TTTGTATACCCACTTTTGTATCTGGCTATTCCCGCCGGTGTATTCTACTATTCTCGAC | 1140 |
| QY | 381 | LeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyrPheAlaGly | 400 |
| DB | 1141 | TTGAAAGACGAGCAGCTTTTGTATCGCTTACTCCGTTCTGTGTCTGTACTTTTCCCGGT | 1200 |
| QY | 401 | ValMetValArgLeuMetLeuThrLeuThrProValIleCysValSerAlaAlaValAla | 420 |
| DB | 1201 | GTTATGGTTAGATGTGTTGACTTTTGACACCACTCATCTGTGTGTCCGCGCGCTCGCA | 1260 |
| QY | 421 | LeuSerLysIlePheAspIleTyrLeuAspPheLysThrSerAspArgLysTyrAlaIle | 440 |
| DB | 1261 | TTGTCCACAGATATTGTGACTCATCTACCTGGATTCAAGACAAAGTACCGCAATACGCCATC | 1320 |
| QY | 441 | LysProIleAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyrLeuTyr | 460 |
| DB | 1321 | AAACCTGGGCACCTACTGGCCAAATGTATGTTTTCGGATCATTCATCTTTATTGTGAT | 1380 |
| QY | 461 | LeuPheValPheHisSerThrTTPValThrArgThrAlaTyrSerSerProSerValVal | 480 |
| DB | 1381 | CTTTTCGCTCTCCATCTCTACTTGGGTAAACAAGAACTGCATACTCTTCTCCTCTCTGTGTT | 1440 |
| QY | 481 | LeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGluAlaTyr | 500 |
| DB | 1441 | TTGCCATCACAAACCCAGATGGTAATTTGGCGTGTATCGACACTTCAGGAAGCGTAC | 1500 |
| QY | 501 | TyrTTPLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTTPTrpAspTyrGly | 520 |
| DB | 1501 | TATTTCGTTAAGAAATGAATCTGTATGAGCACAGTAAGTTTCAGCGCTGGTGGGATTACGGT | 1560 |
| QY | 521 | TyrGlnIleGlyGlyMetAlaAspAcgThrThrLeuValAspAsnAsnThrTrpAsnAsn | 540 |
| DB | 1561 | TACCAAAATTTGGTGCATGGCAGACAGAACCACTTTAGTCCATPAACACACGTGGACAAAT | 1620 |
| QY | 541 | ThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyrGluIle | 560 |
| DB | 1621 | ACTCATCGCCATCGTTGGTGAAGCCATGGCTTCCTCTGAAGAGAAATCTTACGAAAT | 1680 |
| QY | 561 | LeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGly | 580 |
| DB | 1681 | CTAAAGAGCATGATGTCGATTATGCTTGGTCACTTTTGGTGTCTAATTTGGTGTGGT | 1740 |
| QY | 581 | GlyAspAspIleAsnLysPheLeuTTPMetIleArgIleSerGluGlyIleTTPProGlu | 600 |
| DB | 1741 | GGTGATGACATCAACAAATCTTGTGTGATGATCAGAAATTAGCGAGGGAATCTGGCCAGAA | 1800 |
| QY | 601 | GluIleLysGluArgTyrPheTyrThrAlaGluGlyLysTyrArgValAspAlaArgAla | 620 |
| DB | 1801 | GAGATAAAGAGCGTGATTCTTATACCGCAGAGGGAGAATACAGAGTAGATAGTCAAGGGCT | 1860 |
| QY | 621 | SerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPheProGlnLeu | 640 |
| DB | 1861 | TC TGAGACCATGAGAACTCGTACTTTTACAGATGTCCTACAAGATTTTCCCAAAATTA | 1920 |
| QY | 641 | PheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspVal | 660 |
| DB | 1921 | TTCAATGGTGCCAGCCACTGACAGTGCGTCAACAAATGATCACACCATTAGACGTC | 1980 |

| | | | |
|---|------|---|------|
| Qy | 661 | ProProLeuAspTyrPheAspGluValPheThrSerGluAsnTrpMetValArgIleTyr | 680 |
| Db | 1981 | CCACCATTTAGACTACTTCGACGAAGTTTTTTTACTTCCGAAAACTCGATGGTTAGAATATAT | 2040 |
| Qy | 681 | GlnLeuIysLysAspAspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArg | 700 |
| Db | 2041 | CAATTGAAGAGAGTATGATGCCCAAGGTAGAACTTTGAGGGACGTTGGTGAGTTAACGAG | 2100 |
| Qy | 701 | SerSerThrIysThrArgArgSerIleLysArgProGluLeuGlyLeuArgVal | 718 |
| Db | 2101 | TCTTCTACGAAAAACCAGAGGTCCATAAGAGACCTGAATTAGCGTTGAGATC | 2154 |
| RESULT 4 | | | |
| US-10-032-585-6323 | | | |
| ; Sequence 6323, Application US/10032585 | | | |
| ; Publication No. US2003018093A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Terry, Roemer D. | | | |
| ; APPLICANT: Bo, Jiang | | | |
| ; APPLICANT: Charles, Boone | | | |
| ; APPLICANT: Howard, Bussey | | | |
| ; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery | | | |
| ; FILE REFERENCE: 10182-005-999 | | | |
| ; CURRENT APPLICATION NUMBER: US/10/032,585 | | | |
| ; CURRENT FILING DATE: 2001-12-20 | | | |
| ; NUMBER OF SEQ ID NOS: 8000 | | | |
| ; SOFTWARE: PatentIn version 3.1 | | | |
| ; SEQ ID NO 6323 | | | |
| ; LENGTH: 2256 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Candida albicans | | | |
| US-10-032-585-6323 | | | |
| Alignment Scores: | | | |
| Pred. No.: 2,77e-274 Length: 2256 | | | |
| Score: 2738.00 Matches: 509 | | | |
| Percent Similarity: 81.22% Conservative: 88 | | | |
| Best Local Similarity: 69.25% Mismatches: 110 | | | |
| Query Match: 12.80% Indels: 28 | | | |
| DB: 15 Gaps: 3 | | | |
| US-10-028-384-6 (1-718) x US-10-032-585-6323 (1-2256) | | | |

Db 472 TTTATGGGATGCCCCAGGTTATATTTCAAGATCAGTGGCTGCTTCTTATGATTAATGAA 531
Qy 169 AlalleAlalleTherLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThr 188
Db 532 GCAATTTGCATTTCTTATTAATGGAACATTTTATTTCTGGATTAAATCAATGAATG 591
Qy 199 GlySerIleMetHisAlaThrCysAlaAlaLeuPheThrPheMetValSerAlaTrp 208
Db 592 GGTTCAGTTTCTATGCAATTCACAGCATTATTTCTTCTATATGTTAGTCTTGG 651
Qy 209 GlyGlyTrpValPheIleThrAsnLeuLeuPheLeuHisValPheLeuLeuLeuMet 228
Db 652 GGTGGATATGTTTCAATACCAATTTGATTTCCATTCACATGATTTGCTGATTTTCATG 711
Qy 229 GlyArgTrpSerSerLysLeuTrpSerAlaThrThrTrpTrpAlaIleGlyThrVal 248
Db 712 GGTCTTAAATGCAACCTTACACTGCTTACTACATGGTATGCTTGGTACTTGG 771
Qy 249 AlaSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAla 268
Db 772 GCATCAATGCAGATTCATTCGTTGGTGTATTTTACCATTAAGATCAAAATCATATGGCT 831
Qy 269 AlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGln 288
Db 832 GCATPAGGATTTTGGATGTTACAAATAGTGGCTTTGGTATGATTTATTTAAATCAAA 891
Qy 289 IleSerThrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeu 308
Db 892 GTTCCAAACCAACAAATTAATCATCTTGTATGATTTCCATTTCTACTTCTGTTGATTA 951
Qy 309 GlyValValGlyLeuSerAlaLeuThrTrpMetGlyLeuIleAlaProThrPheGlyArg 328
Db 952 GGTATTTGGTGAATTTTGGATTAACAGCAATGGTGGTGGTCTCTTGGACAGTACA 1011
Qy 329 PheTrpSerLeuTrpAspThrAsnTrpAlaLysIleHisIleProIleIleAlaSerVal 348
Db 1012 TTTTATTCCTTATGGATACAAATATGCTCAAGATTCATATTCATTTCTGCTCTGT 1071
Qy 349 SerGlnHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrp 368
Db 1072 TCTGAACATCACTACTGCTGGCCAGCAATCTTTTCTGACTAGTATGCTTATTTGG 1131
Qy 369 LeuPheProAlaGlyValPheLeuLeuLeuLeuLeuLeuLysAspGluHisValPheVal 388
Db 1132 TTTATTCCTGCTGATCTATTTATTTTCTCAAGATTCGAAGATCAACACGCTTTTCAT 1191
Qy 389 IleAlaTrpSerValLeuLeuSerTrpPheAlaGlyValMetValArgLeuMetLeuThr 408
Db 1192 ATCATTTACATGTTATTTGTTCTTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1251
Qy 409 LeuThrProValIleCysValSerAlaValAlaLeuSerLysIlePheAspIleTrp 428
Db 1252 TTGACTCCAGTCATTTGTTGTTGCTGCGACCAATGCTTATCTAAATGTTTGTATGCTAT 1311
Qy 429 LeuAspPhe----- 431
Db 1312 TTGGCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1371
Qy 432 -----LysThrSerAspArgLysTrpAlaIleLysProAlaLeu 445
Db 1372 GAATCCAAAGAGCTCAACCAAAATCAAGTCCAGATTTCCATTT-----GCTGGATAT 1425
Qy 446 LeuAlaLysLeuIleValSerGlySerPheIlePheTrpLeuTrpLeuPheValPheHis 465
Db 1426 TTGTCAAAAGTTTGGTTTACTGACATTTTACATTTTACCTTTTCTACTTTGTTTACAT 1485
Qy 466 SerThrTrpValThrArgThrAlaTrpSerSerProSerValValLeuProSerGlnThr 485
Db 1486 TGTACTTGGGTAACATCGAATGCTTATTCATCCATCAGTGTGTTTGTAGCATCCAGAAC 1545
Qy 486 ProAspGlyLysLeuAlaLeuIleAspPheArgGluAlaTrpTrpLeuArgMet 505
Db 1546 CCAGATGCTCACACATATCATTTGATGATTAAGAGAGGCTTATTTACTGTTAAGATG 1605

Qy 506 AsnSerAspGluAspSerLysValAlaAlaTrpTrpAspTrpGlyTrpGlnIleGlyGly 525
Db 1606 AATPACAGAGAGATGCCAAAGTTATGCTGCTGGTGGATTTATGTTTATCAATCGGGGT 1665
Qy 526 MetAlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIle 545
Db 1666 ATGGCTGATAGAACACACTTGTGATTAACAATACATGGAATAACAACATATTCCTACT 1725
Qy 546 ValGlyLysAlaMetAlaSerProGluLysSerTrpGluIleLeuLysGluHisAsp 565
Db 1726 GTTGGTAAGGCAATGCTTCCCTCAAGATGCTGATGAATTTTGAGACACACAGAT 1785
Qy 566 ValAspTrpValLeuValIlePheGlyLeuIleGlyPheGlyGlyAspAspIleAsn 585
Db 1786 GTTGATTAAGTGTAGTATATTTGGAGGGTATTTGGGTATTTCTGGTATGATATTAAC 1845
Qy 586 LysPheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArg 605
Db 1846 AATTTCTTATGATGCTGAAGATTTGCTGAAGTATCTGGCTGATGAATCAAAAGAAAGA 1905
Qy 606 TyrPheTrpAlaGluGlyGluTrpArgValAspAlaArgAlaSerGluThrMetArg 625
Db 1906 GACTACTTTACTGACCGAGGAGATATTAAGTGGATAAAGATGATCATCTGGCAATGAAG 1965
Qy 626 AsnSerLeuLeuTrpLysMetSerTrpLysAspPheProGlnLeuPheAsnGlyGln 645
Db 1966 AATTTCTTATGATGATTAAGTATCTGATCATGATTTCACTGAATTTGTTGAGGTAGAGAT 2025
Qy 646 AlaThrAspArgValArgGlnMetIleThrProLeuAspValProProLeuAspTrp 665
Db 2026 GGTGTTGTAGTAGTAGTAAGAACCAACAAATCCAGCCATGAAGTACCGAATTAATGTT 2085
Qy 666 PheAspGluValPheThrSerGluAsnTrpMetValArgIleTrpGlnLeuLysLysAsp 685
Db 2086 GTTGAAGAAGCCTTCACATCAGAAAATTTGATTTGAGAAATTTTCAAAAGTTAAAGATTG 2145
Qy 686 AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer----- 702
Db 2146 GATAATGTTGTAGTAGATTTTACATCAAGCTACTGCTTTTGAAGATCATCATCGGCACT 2205
Qy 703 ThrLysThrArgArgSerIleLysArgProGluLeuGlyLeuArg 717
Db 2206 TCCAAAAGAACAGATCCATAAAGACCTAAATTTGAAGTAAGA 2250

RESULT 5
US-10-128-714-7139
Sequence 7139, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroskin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
TITLE OF INVENTION: Methods of Use
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10128,714
CURRENT FILING DATE: 2002-04-23
PRIORITY APPLICATION NUMBER: US 60/285,697
PRIORITY FILING DATE: 2001-04-23
PRIORITY APPLICATION NUMBER: US 60/287,066
PRIORITY FILING DATE: 2001-04-27
PRIORITY APPLICATION NUMBER: US 60/295,890
PRIORITY FILING DATE: 2001-06-05
PRIORITY APPLICATION NUMBER: US 60/303,899
PRIORITY FILING DATE: 2001-07-09
PRIORITY APPLICATION NUMBER: US 60/316,362
PRIORITY FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1

Db 1978 TACAATTTCAACTCTCTCTCCGTCGGCCCAAGCTGTCCAGCCGCTCCGTCGGTCAAAA 2037
 Qy 655 IleThrProLeuAspValProLeuAspTyrPheAspGluValPheThrSerGluAsn 674
 Db 2038 CTT---CCACAGAGAGGCCCTCAGCTCTCTACACTCGAAGAAGCTTTTCACGAGCGAGAAC 2094
 Qy 675 TtpMetValArgIleTyrGlnLeuLysAspAspAlaGlnGlyArgThrLeuArgAsp 694
 Db 2095 TGGATCATCTGATCTATCAAGGTCAAGATCTTGACCAACCTTGGCCGAGACCAACACAG 2154
 Qy 695 ValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLysArg 711
 Db 2155 GCTGTTGCTTCCGACAAA---GGTCTCAAGAAAAAGCGGATCAAAAGAGG 2202

RESULT 6

US-10-320-797-2305
 ; Sequence 2305, Application US/10320797
 ; Publication No. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eroskin, Alexey M.
 ; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
 ; FILE OF INVENTION: METHODS OF USE
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: 60/341,261
 ; PRIOR FILING DATE: 2001-12-17
 ; NUMBER OF SEQ ID NOS: 3361
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 2305
 ; LENGTH: 2466
 ; TYPE: DNA
 ; ORGANISM: Cryptococcus neoformans
 US-10-320-797-2305

Alignment Scores:

Pred. No.: 8,926-229 Length: 2466
 Score: 2301.50 Matches: 435
 Percent Similarity: 73.05% Conservative: 107
 Best Local Similarity: 58.63% Mismatches: 155
 Query Match: 61.19% Indels: 45
 DB: 16 Gaps: 6

US-10-028-384-6 (1-718) x US-10-320-797-2305 (1-2466)

Qy 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerArg 32
 Db 223 GAAAGCTTGTGGCTTCATCATCTCGCGTTAATATGCGGCTCGGATTCGAAATCGA 282
 Qy 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsnTyr 52
 Db 283 TTGTTTGGGTGATCAGATTCGATCTGTATCCACGAATTCACCCCTGGTTCAACTAC 342
 Qy 53 ArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAsp 72
 Db 343 CGAGCCTCGAAAGTCTTGTGTTAAACAAGGGTTTCTACGAGTCTGGAAGCTGGTTGACCCC 402
 Qy 73 ArgThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThr 92
 Db 403 TCGGCTTGGTACCTCTCGGCGAGAACTCTCGGTACCACTCTATCTCGGCTTGATGTC 462
 Qy 93 ThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArg 112
 Db 463 AGCTCTGGACACTGATTGGATCTGCTCTTGGGCA---ATCAATATGCGCGTGACATTCGC 519
 Qy 113 AsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGlu 132
 Db 520 AATGCTGTGCTCTCTTGCACCTGGATTTCTGGATTCGATTCGCTGGCGACTTATCTT 579
 Qy 133 PheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaIleGlyPheIleAlaIle 152
 Db 580 TTCACCACTGAAATGCTTACACCATCAGCTGGTCTATTGGCGCGCGCTTTCATGGCAAT 639

Qy 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIle 172
 Db 640 GTACCCGGATACATCTCTCGATCTCTCGCGGTTCTTATGACAAACGAGCATTCGCATC 699
 Qy 173 ThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMet 192
 Db 700 TTCCTCTTGATGAGCTCTCTACTCTTGGATTAAAGCCGCTCAAAACCGGTAGCTCATTT 759
 Qy 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVal 212
 Db 760 TGGGGTATGATCAGCTGCTTGTCTACGGTGGATGTTGCTGATGGGTGGTTCAGTT 819
 Qy 213 PheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSer 232
 Db 820 TTCATCAACCAATGATTCATTCACGCTTGTCTCATTCATGGCGCATTCAC 879
 Qy 233 SerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGln 252
 Db 880 AACCGGCTTTATACCGCTTACTCTTCCTGGTATGTCATTCGAACTATCGCCTCCATGCAG 939
 Qy 253 IleProPheValIcLysPheLeuProIleArgSerAsnAspHisMetAlaLeuGlyVal 272
 Db 940 GTCCCTTGTGGAGTCTCTCCCATCCGAACCTCTGAGCACAATGGCGGCTTGGGTGT 999
 Qy 273 PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla 292
 Db 1000 TTCGGTCTTGTACAGCTGATCGGATTCGGAAGTGTGCGAGACTCTGTCCTGGCAAG 1059
 Qy 293 LysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyVal 312
 Db 1060 CAATTCAGCTCTCTCAAGCTTTTGTCTGGCGCTTATTCCTGCTCAGTTTGTGTC 1119
 Qy 313 LeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeu 332
 Db 1120 CTGTCATCTTGTGCTTCTCTGGATGATCGCCCTTCTGCTGGAAGATTTTATCTCTT 1179
 Qy 333 TrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGln 352
 Db 1180 TGGGATCTGCGTATCGAAGTCCACATGCCATTTATTCCTCGCTCCGAACACAG 1239
 Qy 353 ProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheProAla 372
 Db 1240 CCACCGCTTGGCGCTCATCTCTTACCTCGAATGCTTATCTCTTTTCCCTGCC 1299
 Qy 373 GlyValPheLeuLeuPheLeuAspLysAspGluHisValPheValIleAlaTyrSer 392
 Db 1300 GGTGCTCTTCTGCTGTTTCAAGGAGCTTCGCGATGAGCAGATCTTCATCATCTTATGCC 1359
 Qy 393 ValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProVal 412
 Db 1360 GTTCTCAGTGCCTATTTTGGCGGTGCTATGTTGCTGCTTATGTTGTCATCAGCCTGTT 1419
 Qy 413 IleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeuAsp 430
 Db 1420 GTCTGTGTTTCTCGCCATTCGCTTCCAACTTCTCGAGGCGTATATGACCCGCTC 1479
 Qy 431 -----PheLysThrSer 434
 Db 1480 ATCCCGAAGCGACGAGGAGCTGGCGAGTCTCAGACGCGAGGTGTCCTCAAGTCCAA 1539
 Qy 435 AspArgLysTyrAlaIleLysProAla -----AlaLeuLeu 446
 Db 1540 GCGAAGAGATGGCGCTGCGCAACGCGCAATAGAGCGGTTCTCTTTCACAGGATATTG 1599
 Qy 447 AlaLeuLeuIleValSerGlySerPhe ----- 455
 Db 1600 AGCGGCAAGTGTCTCGGCACTTTTGGTCTCGACACTCGATTGCTGGTGGTTCCATT 1659
 Qy 456 ---IlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyr 474
 Db 1660 CTCTCTGCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1719

Db 777 GATTGCCATCTTCTGCTTGGTTTCACATCTTCTTATGATCAAGGCTGTCAAAAATGG 836
Qy 189 ySerileMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaThrCl 209
Db 837 GTCATCATGTTGGGAGCGCTGACCGCACATATCTACGGCTACATGGTTCGGCATGGG 896
Qy 209 yGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetG1 229
Db 897 TGGGTATGCTCTTCAATACGAACCTGATCCCCCTGACGATTTTGTCTTCTGTCATGGG 956
Qy 229 yArgTyrSerSerLysLeuTyrSerAlaTyrThrThrTyrAlaIleGlyThrValAl 249
Db 957 TAGATACAGCACTCGCATCTACATAGCTATACACATGATGATGCGCTGGGACTTTGGC 1016
Qy 249 aSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
Db 1017 TAGCATGCAGATTCCTTCGTCGGATTTTGGCTATCCGAACACAGCACCATGTCGCG 1076
Qy 269 aleu-----G 271
Db 1077 CTGGGTAGGTCCTTTCGGAAGTCAAAATGGCGCTTCAGAGCTGACATTTTGCAG 1136
Qy 271 lyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSer 291
Db 1137 GTGCTTCGGCTGCTTTCAGCTTGTGGCTTTCGCGAGTTTGTCCGAGCTTCGTTCCAA 1196
Qy 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValV 311
Db 1197 GCAAGCATCTCCAGAGACTCTGACCGCCATGATCTTCACCTTCGGTCTCGGTTTCG 1256
Qy 311 alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTyrThrGlyArgPheTyrS 331
Db 1257 TTGGGCTAGTGTCTGACTGTGACGGAGTGATCGCTCTCGAGCGCGGATCTACT 1316
Qy 331 erLeuTrpAsnThrAsnTyrAlaLysIleHisIlePhePheIleAlaSerValSerGluH 351
Db 1317 CCTTGTGGACACGTGCTATCCAAAATCCACATCTCCATCTGCTGCTGAGCTCGGAAC 1376
Qy 351 isGlnProValSerTrpProAlaPhePhePheAspThrHisPheLeuIleTrpLeuPheP 371
Db 1377 ACCAGCCACCGCTTGGCAGCGTCTTCTCGATCTGAATCTCTGATCTGGCTTTTC 1436
Qy 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaT 391
Db 1437 CGGAGGTGTCTACATGCTTCGCTGACCTCAAGACGACGATGCTTCGTCATATCT 1496
Qy 391 yrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrP 411
Db 1497 ACTCGGTCTTTCGAGCTACTTTCGCGGTGTATGCTCGACTAATGCTGACCTTGACCC 1556
Qy 411 roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu--- 429
Db 1557 CTATTGTGTGTGCTGCTCTCTGGGCTGTGCTCCATCTTCGACACTATATAGCGA 1616
Qy 430 -----AspPheLysThrSerAsp----- 435
Db 1617 CTACCCCTCCGACACCGCTCTGAAGCGAAAACGAATGAAGACTCGTCTTCACAACCTC 1676
Qy 436 -----ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValS 452
Db 1677 TTCGCTGATGTCGGAGCCCAATGTGATACCTCCCATGTTCTAGATATATAGTA 1736
Qy 452 erGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSerThrTyrValThrArgT 472
Db 1737 CGGCTCTCTGCTGCTTACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
Qy 472 hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaL 492
Db 1797 ATGCATATCTTCT 1856
Qy 492 euIleAspAspPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerL 512
Db 1857 TCATTGAGCATTATCTGAGGCTTACTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTT 1916

Qy 512 yValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrL 532
Db 1917 AATCATGTCATGTTGGGANTATGGTATCAAAATCGTGGCATCGCGACCGCCAACT 1976
Qy 532 euValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaS 552
Db 1977 TGGTTGACACACACACCTGGAACACACCATATTTGCTACGTTTGGTAAGCGCATGAGCT 2036
Qy 552 erProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValI 572
Db 2037 CAGCGAGAGAGTACGATCCCATCTCCGCGAGCATGATGCTGATATCGTCTGCTGG 2096
Qy 572 lePheGlyLeuIleGlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleA 592
Db 2097 TGTTCGTGCTCTGCTAGGTATTCTGGCGATGACATTAACAAATCTTATGATGATGCTC 2156
Qy 592 xGlySerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluG 612
Db 2157 GTATCCCGAAGTATCTGCGCGATGAGGTTAAAGCGCGGACTTCTTACTGACGCG 2216
Qy 612 lyGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeu----- 629
Db 2217 GTCAATATGCTGTCGACGATGAGCGACCCCACTATGCGCAACAGCTTGTATGATGAT 2276
Qy 630 -----Tyr 630
Db 2277 TCCTCTCTCTCTCTTATCTACTGCGTGCATCAGCTAAATTCACAAAATCTAGGTAT 2336
Qy 631 LysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 2337 AAAATGCTCTTATTAACAATTTCAACTCTCTCTCCGTCGGGCCAAGCTGTGACCGGTC 2396
Qy 651 ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAspGluValPhe 670
Db 2397 CGTGGCTCAAACTT---CCACAGAGCGCTCAGCTCTCTACCTCGAAGAGCTTTC 2453
Qy 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLysLysAspAlaGlnGlyArg 690
Db 2454 ACAGCGAGAACTTGGATCATCTGATCTACAGGTCAGGATCTTGAACACCTTGGCCGA 2513
Qy 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgAspSerIleLys 710
Db 2514 GACCAACACCGCTGTTGCTTCGACAAA---GGTCTCAGAAAAAGCGGATACAAAG 2570
Qy 711 Arg 711
Db 2571 AGG 2573

RESULT 8

US-10-128-714-139

Sequence 139, Application US/10128714

Publication No. US20030119013A1

GENERAL INFORMATION:

APPLICANT: Jiang, Bo

APPLICANT: Hu, Weng

APPLICANT: Tishkoff, Daniel

APPLICANT: Zamudio, Carlos

APPLICANT: Eroskin, Alexey M

APPLICANT: Lemieux, Sebastien M

TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

TITLE OF INVENTION: Methods of Use

FILE REFERENCE: 10182-018-999

CURRENT FILING DATE: 2002-04-23

PRIOR FILING DATE: 2001-04-23

PRIOR FILING DATE: 2001-04-27

PRIOR FILING DATE: 2001-04-27

PRIOR FILING DATE: 2001-06-05

PRIOR FILING DATE: 2001-06-05

PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: Patent In version 3.1

; SEQ ID NO 139

; LENGTH: 3969

; TYPE: DNA

; ORGANISM: Aspergillus fumigatus

; US-10-128-714-139

Alignment Scores:

Pred. No.: 1,32e-224 Length: 3969

Score: 2264.50 Matches: 461

Percent Similarity: 66.31% Conservative: 98

Best Local Similarity: 54.69% Mismatches: 135

Query Match: 60.21% Indels: 150

DB: 15 Gaps: 11

US-10-028-384-6 (1-718) x US-10-128-714-139 (1-3969)

QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
 DB 698 CTTCTCGGATATATCTCTGTACCATTCGACGAGCAGGATGCGACCTCTC 757
 QY 35 AlaValIleLys----- 38
 DB 758 AGCGTTATCCG-TAAGTCATCGAGGGAATCAAAGTTGCTAAATTTGTTAATCTATCATCA 816
 QY 39 -----PheGluSerIleIleHisGlu----- 45
 DB 817 AATTGACTGATCATTTCTTTCTACAGGCTTCGAGAGTATCATCCAGGAATGTAAGTAT 876
 QY 45 ----- 45
 DB 877 AGTCAATTGTATPACCTACGGCTTCGGCTTTAAGGACCTCGCTTGGACGAATTGATA 936
 QY 46 -----PheAspProTrpPheAsnTyArgAlaThrLysTyLeuVal 59
 DB 937 CTGATGCTCGGTATATAGTTGACCGGTGTTCAACTTCCGAGCAACAAATATACTTAGTA 996
 QY 60 AsnAsnSerPheTyLysPheLeuAsnTrpPheAsp----- 72
 DB 997 CAGAAATGTTTCTATAGCTTTTGGGATTTGGTTGATGACCGTATGTGGTTCCCAACGCA 1056
 QY 72 ----- 72
 DB 1057 CAAGAGTGCATGAAGCAACAGGTCACTGCTCGGAGTGTCTTAGGCTAATCGGCTCC 1116
 QY 73 -----Arg-ThrTrpTyProLeuGlyArgValThrGlyThrLeuTyProGlyLe 90
 DB 1117 CAACATAGGAACATGGACCTCTGAGAGTGTACCGGTGGCAGTTATATCCCGTCT 1176
 QY 90 uMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs 110
 DB 1177 CATGGTGCAGCGGGGTGATCTACCATATCTTGGCA--TTCCCTTACTATCCCGCTCA 1233
 QY 110 PileArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaTh 130
 DB 1234 TATTCGACATCTGGTCTACTGCGCCAGGATCTCCGGCTCGATGCTATGCGCAAT 1293
 QY 130 rTyrgluPheThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPh 149
 DB 1294 GTACTTCTGACATCGGAGATGCTCTCTCGCCATCTGCGGTCTTCTTGCAGCAGCTTT 1353
 QY 149 eIleAlaIleValProGlyTyIleSerArgSerValAlaGlySerTyAspAsnGluAl 169
 DB 1354 CATGGGAATCGCCCTGGTTTACATCTCCCGATCAGTGTGGAAGCTACGATAACGAAGC 1413
 QY 169 aIleAlaIleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGl 189
 DB 1414 GATTGCCATCTTCTGCTGTGTTCACATTTCTTCTATGGATCAAGGCTGCAAAAATGG 1473
 QY 189 ySerIleMetHisAlaThrCysAlaAlaLeuPheTyPheTyMetValSerAlaTrpGl 209

DB 1474 GTCTATCATGTGGGAGCGCTGACCGCACTATTCTACGGCTACATGGTCTCGCATGGGG 1533
 QY 209 YGlyTy-ValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGl 229
 DB 1534 TGGGTATGTCTTCATTACGAACCTGATCCCTGCGACGTTTTTGTCTTCTGTGATGGG 1593
 QY 229 YArgTySerSerLysLeuTySerAlaTyThrThrTyTrpTyAlaIleGlyThrValAl 249
 DB 1594 TAGATACAGCACTCGCATCTACATTAGCTATACCATGGTATGGCTGGGACATTGGC 1653
 QY 249 aSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
 DB 1654 TAGCATGCGATTCCTTCGTCGATTTTTCGCTATCCGAAACAGCAGCACCATGTCCGC 1713
 QY 269 aLeu-----G 271
 DB 1714 CTTGGGTAGTCTCTGCTTCCGAAAGTTCAAATGGCGGTTCAGGAGCTGACATTTTGGNG 1773
 QY 271 lYValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerT 291
 DB 1774 GTGCTTCGGCTCTGCTTTCAGCTTGTGGCTTTCGCGAGTTTTCGCGAGCTTCGTTCCAA 1833
 QY 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuLeuValLeuGlyValV 311
 DB 1834 GCAAGCAGTTCGAGAGCTTCTGACCGCATGATCTTCATCACCCTTCGCTTCGGTTTCG 1893
 QY 311 alGlyLeuSerAlaLeuThrTy-MetGlyLeuIleAlaProTrpThrGlyTyArgPheTyS 331
 DB 1894 TTGGCTAGTTGTTCTGCTGTGACGGAGTGTGCTCTCTGGAGCGCGCATCTACT 1953
 QY 331 erLeuTrpAspThrAsnTyAlaIleHisIleProIleIleAlaSerValSerGluH 351
 DB 1954 CTTGTGGACACTGGCTATGCCAAATCCAAATCCCATCTCCATCTGCTCAGTCTGGAAAC 2013
 QY 351 isGlnProValSerTrpProAlaPhePhePhePheThrHisPheLeuIleTrpLeuPheP 371
 DB 2014 ACCAGCCACCGCTTGGCCAGCTTCTTCTTCGATCTGAACCTTCCTGATCTGGCTTTCC 2073
 QY 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla 391
 DB 2074 CGCGAGGTCTATCTATGTGTTCCGTGACCTCAAGACGAGCATGTCTTCGTCATTAICT 2133
 QY 391 yrSerValLeuCysSerTyPheAlaGlyValMetValArgLeuMetLeuThrLeuThrp 411
 DB 2134 ACTCGTCTTTCGAGGTACTTTCGCGTGTATTATGTCGCACTAATGCTGACCTTGACCC 2193
 QY 411 roValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyLeu---- 429
 DB 2194 CTATTGTGTGTGTCGCTGCTGCGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2253
 QY 430 -----AspPheLysThrSerAsp----- 435
 DB 2254 CTACCTCTCCGACACCGCTCTGAAGCGAAACGAATGAAGACTCGTCTTCACACCTTC 2313
 QY 436 -----ArgLysTy-AlaIleLysProAlaAlaLeuLeuAlaLysLeuValS 452
 DB 2314 TTCGCTCAGTTTCGGAAGCCCAATGTTGGAATCACCTCCCATGTTCTTAAGATTATGTA 2373
 QY 452 erGlySerPheIlePheTyLeuTyLeuPheValPheHisSerThrTrpValThrArgT 472
 DB 2374 CGCGCTCTGTGTCTGCTTACCTGCTGTTTGTGCGCACTGCACTCGCTGGTTTACATCGA 2433
 QY 472 hrAlaTySerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaL 492
 DB 2434 ATGCATATCTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2493
 QY 492 eIleAspAspPheArgGluAlaTyTrpTrpLeuArgMetAsnSerAspGluAspSerL 512
 DB 2494 TCATTGAGGATATTCGTGAGCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2553
 QY 512 ysValAlaAlaAlaTrpAspTyGlyTyGlnIleGlyMetAlaAspArgThrThrL 532

Db 2554 AAATCATGTCATGTTGGGATTTATGGGTATCAAAATCGGTGGCATGGCGGACCCCAACCT 2613
QY 532 euValAspAsnThrTrpAsnAsnThrHisLeuValLeuValGlyLysAlaMetAla 552
Db 2614 TGGTGTGACAAACACCTGGAACACCACTATTTCTAGGTTGGTAAAGGGGATGACT 2673
QY 552 erProGluGluLysSerTyrGluLeuLeuLysGluHisAspValAspTyrValLeuVal 572
Db 2674 CACGCGAGGAGTCACTACCCATCTCCGCCAGCATGATGTCGATTACGTTGGTGG 2733
QY 572 lePheGlyGlyLeuLeuGlyPheGlyAspLeuLysPheLeuTrpMetIleA 592
Db 2734 TGTGGGTTGGTCTCTAGGTTATTTCTGGCGATGACATTAACAAATTTCTATGGTGGTCC 2793
QY 592 gileSerGluGlyLeuTrpProGluGluLeuLysGluArgTyrPheTyrThrAlaGluG 612
Db 2794 GTATCGCCGAGGATCTCGGCCGATGAGGTTAAAGAGCGGACTTCTTTACTGCACGCG 2853
QY 612 yGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeu 629
Db 2854 GTGAATATCGTGTGACGATGAGCGACGCCCACTATGCGCAACAGCTTGATGGAT 2913
QY 630 -----Tyr 630
Db 2914 TCCCTCTTCTCCCTCTATCTACTGCGCTGCATCAGCTTAATTCACAAATCTAGGTAT 2973
QY 631 LysMetSerTyrLysAspPheProGluLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
Db 2974 AAAATGCTTTATTAACAATTTCAACTCTCTTCCCGTGGCGCAAGCTGTGACCGCGTC 3033
QY 651 ArgGlnGlnMetIleThrProLeuAspValProLeuAspTyrPheAspGluValPhe 670
Db 3034 CGTGGGTCAAAACTT---CCACAGAGGCGCTCTCTACACTCGAAGAGCTTTC 3090
QY 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArg 690
Db 3091 ACGAGCGAGACTGGATTCATTCGATCTCTCAAGGTCAAGATCTTCAACACCTTGGCCGA 3150
QY 691 ThrLeuArgAspValGlyGluLeuThrArgSerSerThrLysThrArgArgSerIleLys 710
Db 3151 GACCACCAACGAGGTGTGGCTTCGACAAA---GGTCTCAAGAAAAGCGGAGTCAAAAG 3207
QY 711 Arg 711
Db 3208 AGG 3210

RESULT 9
US-10-128-714-5139
; Sequence 5139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5139
; LENGTH: 4603
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5139
Alignment Scores:
Pred. No.: 1-66e-224 Length: 4603
Score: 2264.50 Matches: 461
Percent Similarity: 66.31% Conservative: 98
Best Local Similarity: 54.63% Mismatches: 135
Query Match: 60.21% Indels: 150
DB: 15 Gaps: 11
US-10-028-384-6 (1-718) x US-10-128-714-5139 (1-4603)
QY 15 IleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeuPhe 34
Db 1061 CTTCTCGGATTTATTTCTGTGTACATTGTCAGAGCAGCGATTGCCAGTGCACCTTC 1120
QY 35 AlaValIleLys----- 38
Db 1121 AGCGTTATCCG-TAAGTTCATCGAGGAAATCAAAGTTGCTAAATTTGTTAATCTATCATCA 1179
QY 39 -----PheGluSerIleIleHisGlu----- 45
Db 1180 AATTGACTGATCATTCTTTCTACCAGGCTTCAGAGATATCATCCAGAAATGTAAGTAT 1239
QY 45 ----- 45
Db 1240 AGGTCAATTTATATACCTACGCGCTTCGCGCTTTAAACGACCTCGCTTGGACGAATTGATA 1299
QY 46 -----PheAspProIlePheAsnTyrArgAlaThrLysTyrLeuVal 59
Db 1300 CTGATGCTCCGTGATATAGTTGACCCGTTGGTTCACCTTCGAGCAACAAATATCTTAGTA 1359
QY 60 AsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAsp----- 72
Db 1360 CAGAAATGGTTTCTATAGCTTTTGGGATTTGTTGATGACCGTATGTGGTTCCCAACCGCA 1419
QY 72 ----- 72
Db 1420 CAAGATGATGAAGCAACAGGTCACTGTCTCGAGTGTCTTAGCTAATCGGCGTCC 1479
QY 73 -----Arg-ThrTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLe 90
Db 1480 CAACATAGGAACATGGCATCTCTGGGACGTGTCCCGGTGGCAGCTATATATCCCGTCT 1539
QY 90 uMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAs 110
Db 1540 CATGGTGACGAGCGGCGGTGATCTACCATATCTTGGGA---TTCCTTACTATCCCGTCCA 1596
QY 110 pIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAla 130
Db 1597 TATTCGAACATCTCGCTCTACTGGCGCCAGGATTTCCGCGCTGCTGCTGATGGCAAT 1656
QY 130 rTyrGluPheThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaGlyP 149
Db 1657 GTACTTGTGATCATCGAGATGTCTCTTCGCTTCGAGTCTTCGAGGTCTTCTTCGACAGCTTT 1716
QY 149 eIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAl 159
Db 1717 CATGGGAATCGGCCCTGTTTACATCTCCCGATCATGTTGCTGGAAAGCTTACCATCAACGAAGC 1776
QY 169 aIleAlaIleThrLeuLeuMetValThrPheMetPheTyrIleLysAlaGlnLysThrGl 189
Db 1777 GATTGCAATCTTCTGCTGTGTTCCACATTTCTTCTATGGATCAAGCTGTCAAAATGG 1836
QY 189 ySerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGl 209
Db 1837 GTCTATCATGTGGGAGCGGTGACCGCACTATTTACGGCTACATGTGTGCGATGGG 1896

QY 209 yGlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGl 229
DB 1897 TGGGTATGCTTCAATACGAACCTGATCCCTGACGCTTTTGTGCTTCTGTGATGGG 1956
QY 229 yAqTyrSerSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAl 249
DB 1957 TAGNACAGCAGCTCGCATCTACATTAAGCTATACCACTAGGTATGGCTGGGATTTGGC 2016
QY 249 aSerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAl 269
DB 2017 TAGCATGAGATTCCTTCGCTCGGATTTTGGCTATCCGAACACAGCAGCATGTCCGC 2076
QY 269 aLeu-----G 271
DB 2077 CTGGGTAGTCTTCCTTCGCAAGTTCAAATGCGCTTCAGGAGCTGACATTTTCAG 2136
QY 271 lyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerT 291
DB 2137 GTGCTTCGGCCCTGCTGAGCTTGTGGCTTCGCGAGTTGTTCGACAGCTTCGTTCCAA 2196
QY 291 hrAlaLysPheLysValIleMetMetValSerLeuPheLeuIleValLeuGlyValV 311
DB 2197 GCAAGCAGATTCAGAGACTTCGACCGCCATGATCTTCATCAGCTTCGGTTCG 2256
QY 311 alGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrS 331
DB 2257 TTGGCTAGTGTTCGACTGTGACGGAGTGTGCTCTTCGAGCGGCGGATTTCTACT 2316
QY 331 erLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluH 351
DB 2317 CTTGTGGACACTGGCTATGCGAAATCCATTCCTCATCTGCTTCAGTTCGGAAC 2376
QY 351 isGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheP 371
DB 2377 ACAGGCCACCGTTCGGCAGGCTTCTTCGATCTGCACTTCCTGATCTGGCTTTTC 2436
QY 371 roAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAla 391
DB 2437 CGCAGGTGTCTACAGTGTCTTCGCTGACCTCAAGAGCAGCATGTCTTCGCTATCT 2496
QY 391 yrSerValLeuCySsrTyrPheAlaGlyValMetValArgLeuMetLeuTrpLeuThrp 411
DB 2497 ACTCGTCTTCGAGCTACTTCGCGGTGTATGCTCGACTAATGCTGACCTTGACCC 2556
QY 411 roValIleCySsrValSerAlaValAlaLeuSerLysIlePheAspIleTyrLeu---- 429
DB 2557 CTATTGTGTGTGCTGCTGCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2616
QY 430 -----AspPheLysThrSerAsp----- 435
DB 2617 CTACCTCCGACACAGGCTGTGAAGCGGAACAGATGAAGCTGCTGCTTCACAACTC 2676
QY 436 -----ArgLysTyrAlaIleLysProAlaLeuAlaLeuLysLeuIleValS 452
DB 2677 TTCGCTCAGTTCGGAAGCCCAATGTGGGAATCACCTCCCATGTTCTTAAGATTATGTA 2736
QY 452 erGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgT 472
DB 2737 CGCGCTGTGTGCTGCTACCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2796
QY 472 hrAlaTyrSerSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAla 492
DB 2797 ATGCATCT 2856
QY 492 euIleAspAspPheAspGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerL 512
DB 2857 TCATTGACGATTATCGTAGGCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2916
QY 512 ySValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGlyLeuMetAlaAspArgThrL 532
DB 2917 AAATCATGTGATGGGTATGGGTATCAAAATCGGTGCTGCTGCTGCTGCTGCTGCTG 2976
QY 532 euValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAla 552

DB 2977 TGGTGTGACAAACACCTTGGAAACACCCATATTTGTCGGTGTGTAAGGCGATGAGCT 3036
QY 552 erProGluGluLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValI 572
DB 3037 CACGCGAGAGTTCAGTACCCATCTCCGCGAGCATGATGTCGATACGTCGTGTGG 3096
QY 572 lePheGlyGlyLeuIleGlyPheGlyGlyAspIleLeuLysPheLeuTrpMetIleA 592
DB 3097 TGTTCGGTGGTCTCTAGGTATTTCTGGCGATGACATAACAAATCTTATGATGGTCC 3156
QY 592 yGlySerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluG 612
DB 3157 GTATCGCGAAGGTATCTGGCCGATGAGGTAAAGAGGGGACTTCTTTACTGACGCG 3216
QY 612 lyGluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeu----- 629
DB 3217 GTGAATATCGTGTGAGATGGAGCGACCCCACTATGGCGAACAGCTTGATGTATGAT 3276
QY 630 -----Tyr 630
DB 3277 TCCCTCTTCCTCCCTCTTATCTACTGCGCTGCATCAGCTTAATTCACAAATCTAGGTAT 3336
QY 631 LysMetSerTyrLysAspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgVal 650
DB 3337 AAATGTCTTATTAACAATTTCACTCTCTTCCGTCGGGCGCAAGCTGTTCGACCGCTC 3396
QY 651 ArgGlnGlnMetIleThrProLeuAspValProLeuAspTyrPheAspGluValPhe 670
DB 3397 CGTGGTCAAACTT---CCACAGAAAGGCTCTCTCTACACTCGAAGAAGCTTTC 3453
QY 671 ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArg 690
DB 3454 AGACGAGAACTGGATCTGATCTGATCTCAAGCTCAAGATCTTCAACCTTGCCCGA 3513
QY 691 ThrLeuArgAspValGlyLeuLeuThrArgSerSerThrLysThrArgSerIleLys 710
DB 3514 GACCACACAGGCTGTGCTTCGACAAA---GGTCTCAAGAAAAAGCGGAGTACAAAG 3570
QY 711 Arg 711
DB 3571 AGG 3573
RESULT 10
US-10-128-714-2139
; Sequence 2139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wendi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Ershkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2139
; LENGTH: 1848

! TYPE: DNA
! ORGANISM: Aspergillus fumigatus
US-10-128-714-2139

Alignment Scores: 6.01e-207 Length: 1848
Pred. No.: 2090.00 Matches: 394
Score: 79.53% Conservative: 80
Percent Similarity: 66.11% Mismatches: 93
Best Local Similarity: 55.57% Indels: 29
Query Match: 15 Gaps: 5

US-10-028-384-6 (1-718) x US-10-128-714-2139 (1-1848)

Qy 63 PheTyrLysPheLeuAsnTrpPheAspAsp-----Arg-ThrTr 75
Db 6 TTCTATAGCTTTTGGGATGGTTTGTAGACCGCTAATCGCGCTCCCAACATAGGAACATG 65
Qy 75 pTyrProLeuGlyArgValThrGlyThrLeuTyrProGlyLeuMetThrSerAl 95
Db 66 GCATCCTCTGGACGCTGTCCCGGTGGCAGCTTATATCCCGCTCTCATGTGACGAGCG 125
Qy 95 aPheLeuTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAsnValGly 115
Db 126 CGTGATCTACCATATCTTGGCA---TTCTTACTATCCCGCTCGATATCGCAACATCG 182
Qy 115 sValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyGluPheThr 135
Db 183 CGTCTCTACTGGCGCGCAGGATCTCCGCGCTGACCTGCAATGGCAATGACTTGTGACATC 242
Qy 135 sGluLeuLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheIleAlaValPr 154
Db 243 CGAGATGCTCTCTCGCATCTGAGGCTCTCTGAGAGCTTTTCATGGATCGCCCG 302
Qy 154 oGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleThrLe 174
Db 303 TGGTTACATCTCCCGATCAGTTGCTGGAAGCTAGCATACGAAGCGATTCCTCTTCT 362
Qy 174 uLeuMetValThrPheMetPheTyrPheLysAlaGlyThrGlySerIleMetHisAl 194
Db 363 GTTGTGTTCATCTCTTATGATGATCAGGCTGTCAAAATGGGTCTATCATGTGGGG 422
Qy 194 aThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheI 214
Db 423 AGCGTGTACCGGACATATCTACGGCTACATGGTGTGCGCATGGGGTGGGTATGCTTCAT 482
Qy 214 eThrAsnLeuLeuProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSerSer 234
Db 483 TAGCAACCTGATCCCTCTGACGCTTTTGTCTCTGTCATGGGTAGATACAGCACTCG 542
Qy 234 sLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIlePr 254
Db 543 CATCTACATATTAGTATACCATCATGATGGCTGGGCACTTTGGCTAGCATGCGATTC 602
Qy 254 oPheValGlyPheLeuProIleAsnSerAsnAspHisMetAlaAlaLeuGlyValPheG 274
Db 603 CTTCTGCGATTTTGTCTATCCGAACAGCAGCCACATGTCCGCTTGGGTCTCTGG 662
Qy 274 yLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysP 294
Db 663 CTTGCTTACGCTTGGCGAGTTTGTCCGAGCTTGTCCGAGCTTGTCCCAAGCAGGAT 722
Qy 294 eLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValValGlyLeuSe 314
Db 723 CCAGAGACTCTGACCGGATGATGATCTTCATCACCTTCGGTCTCGGTTGCTGGGCTAGT 782
Qy 314 rAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpAs 334
Db 783 TGTTCGACTGTGACGGAGATGATGCTCTCTTGGAGCGCGGATTTCTCTCTCTTGGGA 842
Qy 334 pThrAsnTyrAlaLysIleHisIleProIleAlaSerValSerGluHisGlnProVa 354
Db 843 CACTGGCTATGCCAAATCAATCTCCCATCATTCCTGCTGAGTCTCGGAACACGAGCCAC 902

Qy 354 lSerTrpProAlaPhePhePheAspThrHisPheLeuIleTrpLeuPheProAlaGlyVa 374
Db 903 CGCTTGGCCAGCGTTCTTCTTCGATCTGAACCTCTGATCTGGCTTTCGCGCAGGTGT 962
Qy 374 lPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerVal 394
Db 963 CTACATGCTCTCCGTGACCTCAAGGAGCAGCATGCTTCTGTCATTTACTTCTCTCTCT 1022
Qy 394 uCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValleCy 414
Db 1023 TGGAGCTACTTCCGCGGTGTATGTCGCGACTAATGTCGACTGACCTTACCCCTATTGTGT 1082
Qy 414 sValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu----- 429
Db 1083 TGTTCGCGTGTCTGCGGCTGTGTCCTCATCTCGACACCTATATGGCGACTACCTCC 1142
Qy 430 -----AspPheLysThrSerAsp----- 435
Db 1143 GACACCGAGCTCTGAGCGGAAACGAATGAAGACTCTGCTTTCACAACTCTTCGCTCAGT 1202
Qy 436 -ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPh 455
Db 1203 TCGGAAGCCCAATGTGGAAATCACCTCCATGTTTCTTAAGATTATAGTGAGCGCTGT 1262
Qy 455 eLlePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSe 475
Db 1263 TGTGCTACTACCTGCTCTGTTTGTGGCACTGACCTGGGTACATCGAATGCAATCTC 1322
Qy 475 rSerProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAs 495
Db 1323 TTCTCTCTCGGTTGCTTGGCTAGTCCGATCCCTGACGGAAGCAATACATCATTCACCGA 1382
Qy 495 pPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAl 515
Db 1383 TTATCTGAGGCTTACTACTGCTTCTGCAAGATCTCTCAGAACGCCAAATCATGTC 1442
Qy 515 aTrpTrpAspTyrGlyTyrGlnIleGlyLysAlaAspArgThrThrLeuValAspAs 535
Db 1443 ATGGTGGATTATGGTATCAATCGGTGGCATCGCGAGCCGCCAACCTTGTGTGACAA 1502
Qy 535 nAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu 555
Db 1503 CAACACCTGGAAACACCCCATATGTCTACGTTGGTAAAGCGATGATGATCGCGAGGA 1562
Qy 555 uLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGly 575
Db 1563 AGTCAGCTACCCCACTCTCCCGCAGCATGATGCTAGTTCGTCGTGGTGTGGTGG 1622
Qy 575 yLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerG 595
Db 1623 TCTGTAGTATTATCTGGCGATGACATTAACAATTTCTTATGATGTCGCTGATCCGCG 1682
Qy 595 uGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrAr 615
Db 1683 AGTATCTGGCCGATGAGTTAAAGAGCGGACTTCTTACTGCACCGCGGTGAATATCG 1742
Qy 615 gValLeuAlaAlaSerGluThrMetArgAsnSerLeuLeuTyr 630
Db 1743 TGTGCGATGAGGAGGACCCCACTATCGCAACAGCTTGTATGTAT 1788

RESULT 11

US-10-128-714-1139
; Sequence 1139, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Weng
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Frostkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10/128,714
 ; PRIOR FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/297,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1139
 ; LENGTH: 1969
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 ; US-10-128-714-1139

Alignment Scores:

Pred. No.: 6,59e-202 Length: 1969
 Score: 2042.00 Matches: 394
 Percent Similarity: 74.41% Conservative: 80
 Best Local Similarity: 61.85% Mismatches: 93
 Query Match: 54.29% Indels: 70
 DB: 15 Gaps: 6

US-10-028-384-6 (1-718) x US-10-128-714-1139 (1-1969)

Qy 63 PheTyrLysPheLeuAsnTrpPheAsp 72
 Db 6 TTCTATAGCTTTTGGGATTGGTTGATGACCGTATATGGTGTCCACGCAAGAGTGC 65
 Qy 73 -----Arg 73
 Db 66 ATGAAGCAACAGGTCACTGCTGGAGTGTCTTAGGCTAATCGCGTCCCAACATAGG 125
 Qy 74 -ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThr 93
 Db 126 AACATGGCATCTCTGGGACGTGTACCGGTGGACGTTATATCCGGTCTCATGGTGC 185
 Qy 93 rSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAs 113
 Db 186 GAGCGGCGTCACTACCATCTTGGA--TTCTTACTATCCCGTGGATATTCGAA 242
 Qy 113 nValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPh 133
 Db 243 CATCTGGCTCTACTGGCGCAGATCTCCGGCTGACTGATTTGGCAATGTACTTGGCT 302
 Qy 133 eThrLysGluIleLysAspAla---SerAlaGlyLeuLeuAlaAlaGlyPheIleAla 152
 Db 303 GACATCGAGATGTCCTTCGCCATCTCCAGGTCTTCTTGGCAGCAGCTTTCATGGGAT 362
 Qy 152 eValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 172
 Db 363 CGCCCTGGTGTACTCTCCGATCTCTGGAGTGTACGATACGATACGAGGATGGCAT 422
 Qy 172 eThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnTyrGlySerIleMe 192
 Db 423 CTCTCTGCTGTGTTCACATCTCTTCTATGGATCAAGGCTGTCAAAATGGGTCTATCAT 482
 Qy 192 tHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrVa 212
 Db 483 GTGGGAGCGCTGACCGCATATTCTACGGCTACATGGTGTGGGATGGGTGGTNGT 542
 Qy 212 lPheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTyrSe 232
 Db 543 CTTCATTACGAACCTGATCCCTGACCGTCTTCTCTCTGTCATGGGTAGATACAG 602
 Qy 232 rSerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetG 252

Db 603 CACTCGCATCTACATTAGCTATACCATGGTATGGCTGGGACTTTGGCTAGCATGCA 662
 Qy 252 nIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeu--- 270
 Db 663 GATTCCCTTCGTGGATTTTGGCTATCCGAACAGGACCATATGTCGCTTGGGTAG 722
 Qy 271 -----GlyValPheG 274
 Db 723 GTCCTTGTCTTTCGCAAGTTCAAATGGCGCTTCAGGAGCTGACATTTTGCAGGTGTCTTCG 782
 Qy 274 lVLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLysP 294
 Db 783 GCTGCTTCAGCTTGTGGCTTTCGCCAGTTCGCGAGCTTCGCGAGCTTCGCAAGCACT 842
 Qy 294 lLeValIleMetMetValSerLeuPheLeuLeuValLeuGlyValValGlyLeuS 314
 Db 843 TCCAGAGACTTCTGACCGCCATGATCTTCATCACTTCGCTTCGCTTCGCTTCGCTAG 902
 Qy 314 eAlaAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyrSerLeuTrpA 334
 Db 903 TTGTTCTGACTGTGACGGGAGTGTCTCTTCGATCTGAATCTCTGATCTGCTTTTCCGCGCAGGTG 962
 Qy 334 sPThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProV 354
 Db 963 ACATGGCTATGCCCAAAATCCCATTCCTTCATTCCTTCAGTCTCGGACACACAGCCCA 1022
 Qy 354 alSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPheProAlaGlyV 374
 Db 1023 CGCTTGGCCAGCGTCTCTTCGATCTGAATCTCTGATCTGCTTTTCCGCGCAGGTG 1082
 Qy 374 alPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerVal 394
 Db 1083 TCTCATGCTGCTTCGCGACCTCAAGACGAGCATGCTTCGTCATATCTACTCGGTGC 1142
 Qy 394 euCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIleC 414
 Db 1143 TTGGGAGCTACTTCGCGGTGTTATGTTCGAGCTTAATGTGCTGACCTTGAACCTTATTGTGT 1202
 Qy 414 yValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu----- 429
 Db 1203 GGTTCGCGCTGCTCTGCGCTGTCGTCATTCGACACCTATATATGGGACATACCTCTCC 1262
 Qy 430 -----AspPheLysThrSerAsp----- 435
 Db 1263 CGACACAGCGCTCTGAAGCGAAACGAATGAAGACTCGTCTTCAACACCTCTTCGCTCAG 1322
 Qy 436 --ArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerP 455
 Db 1323 TTGGGAGCCCAATGTTGGAAATCACCTCCCATGTTTCTAAGATTATAGACGGGCTCTG 1382
 Qy 455 llePheTyrLeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrS 475
 Db 1383 TTTCGCTACTGCTCTGCTGTTTGTTCGCACTGCACCTGGGTATACATGGAATGATCAT 1442
 Qy 475 rSerProSerValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspA 495
 Db 1443 CTCTTCCTTCGTTGTTTGGTATGTCGATGCTGACGGAACCAATACATCATTTGACG 1502
 Qy 495 sPpPheArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAla 515
 Db 1503 ATTATCGTGGGCTTACTTACTGCTCAGAAATATCTCTCAGAAACGCAAAATCATGT 1562
 Qy 515 lATrTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAspA 535
 Db 1563 CATGGTGGATATGGGTATCAAAATCGTGGCATGGCGACCCGCCCAACCTTGGTTGACA 1622
 Qy 535 snAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluG 555
 Db 1623 ACAACACCTGGAACAACACCATATTCCTACGGTTGGTAAAGGGGATGAGCTCACGGGAG 1682
 Qy 555 lLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyG 575

Db 1683 AAGTCAGTACCCCACTCTCCGCCAGCATGATGTCGATTAGCGTGTGGTGGTTCGGTG 1742
 Qy 575 lyleuileedlyPheGlyAspAspIleAsnLysPheLeuTrpMetIleAArgIleSerG 595
 Db 1743 GTCTGCTAGTATTCTTGGCGATGACATTAACAAATCTTATGGATGGTCCGTATCGCG 1802
 Qy 595 luegylleTrpProGluGluIleLysGluArgTyPheTyThrAlaGluGlyGluTyA 615
 Db 1803 AAGGTATCGGCCCATGAGGTAAAGACGGGACTTCTTACTGACCGCGTGAATATC 1862
 Qy 615 rGValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyR 630
 Db 1863 GTGTCGACGATGGAGCGACCCCACTATATGCGCAACAGCTTGATGAT 1909

RESULT 12

US-10-320-797-1305
 ; Sequence 1305, Application US/10320797
 ; Publication NO. US20040014955A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eroskin, Alexey M.
 ; APPLICANT: Zamudio, Carlos
 ; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
 ; FILE REFERENCE: 10182-021-999
 ; CURRENT APPLICATION NUMBER: US/10/320,797
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: 60/341,261
 ; PRIOR FILING DATE: 2001-12-17
 ; NUMBER OF SEQ ID NOS: 3361
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1305
 ; LENGTH: 2882
 ; TYPE: DNA
 ; ORGANISM: Cryptococcus neoformans
 US-10-320-797-1305

Alignment Scores:

Pred. No.: 1.16e-200 Length: 2882
 Score: 2032.50 Matches: 433
 Percent Similarity: 60.86% Conservative: 105
 Best Local Similarity: 48.98% Mismatches: 159
 Query Match: 54.04% Indels: 191
 DB: 16 Gaps: 14

US-10-028-384-6 (1-718) x US-10-320-797-1305 (1-2882)

Qy 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerArg 32
 Db 223 GAAAGCTTGTTCGGCTTCATCATCTCGCGTTAATATGCGGTGCTCGGATTGGAATCGA 282
 Qy 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPhe----- 46
 Db 283 TTGTTTGGGTGATCAGATTGGAATGTGTATCTATCCACGAATTTGAGTCAAACTATCGCG 342
 Qy 47 -----AspProTrpPheAsnTyR-- 52
 Db 343 AGGATTCCTCAATGCTAAGCTGTGTGCTTCATTTCAGTGACCCCTCGGTCAACTA-GT 401
 Qy 53 -----ArgAlaThrIleTyR 57
 Db 402 AAGGCTATTCTCGAGGTCAAACTCTCCCTACTGATTTATTCCTAGCGAGCCCTCGAAAGT 461
 Qy 57 rIeuValAsnAsnSerPheTyLysPheLeuAsnTrpPheAspArgThrTrpTyR 77
 Db 462 TCTTGTAAACAGGGTTCTTACGAGTTCTCGAATCGTTTGGACCCCTCGCGTGTGTACCC 521
 Qy 77 oleuGlyArgValThrGlyGlyThrLeuTyProGlyLeuMetThrThrSerAlaPheI 97
 Db 522 TCTCGGACAGCTGTCGGTACCAAGCTCTATCTCGGTGATGATGTCACGCTCTGGACTGAT 581
 Qy 97 eTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsnValCysVal 117
 Db 582 TTGGCATGCTCTTCGGGCA---ATCAATATGCGCGTGACATTCGCAATGTCTGTGTCT 638

Qy 117 uPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyR----- 131
 Db 639 CTTTGACACCTGGATTTTCTGGATTGACTGCTGGCGCACTTATCTGTCACTGTAACAAAT 698
 Qy 132 -----GluPheThrLysGluIleLysAs 139
 Db 699 TATCTGATCATTCGATACACATAACATGAGAGCGTTGTAGTTTACCACTGAAATGCTAC 758
 Qy 139 pAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleValProGlyTyRLeuSer 159
 Db 759 ACCATCAGCTGCTATTGGCGCGCTTTCATTGGCATTGTACCGGATACATCTCTCG 818
 Qy 159 gSerValAlaGlySerTyRAspAsnGluAlaIleAlaIleThrLeuLeuMetValThrPh 179
 Db 819 ATCTGTGCGCGGTCTTATGACACAGAGCCATTGCCATCTCTCTTGTAGAGCTCCT 878
 Qy 179 eMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAla 199
 Db 879 CTACTCTTGGATTAAGCGCGTCAAAACCGTAGCTCAATTTGGGGTATGATCACTGCTT 938
 Qy 199 uPheTyRThrMetValSerAlaTrpGlyGlyTyRValPheIleThrAsnLeu----- 217
 Db 939 GTTCTACGGGTGGATGGTGTGTCATGGGTGGTGTACGTTTTCATCACCACAG-TATGT 997
 Qy 218 -----IleProLeuHisValPh 223
 Db 998 CGCTCGGCCCTCAATTGAATGTCTGTTTACTCTTTTGCAGTGATTCATTTGCACGCTT 1057
 Qy 223 eLeuLeuIleLeuMetGlyAArgTyRSerSerLysLeuTyRAlaTyRThrTrpTy 243
 Db 1058 TGTCTCATTTGATGCGGCGGTTCAACACCGGCTTTATACCGCTTACTCTTCTCGTGA 1117
 Qy 243 rAlaIleGlyThrValAlaSerMetGlnIleProPheValGlyPheLeuProIleArg 263
 Db 1118 TGTCAATGGAATCTCGGCTCCATGACAGTCCCTTTGTGGAGTCTCTCCCAATCGAAC 1177
 Qy 263 rAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAs 283
 Db 1178 CTCTGAGCACATGCGCGGCTTGGGTGTTTTCGGTCTTGTACAGTGATCGGATTCGTCA 1237
 Qy 283 pPheValLysGlyGlnIleSerThrAlaLysPheValIleMetMetValSerLeuPh 303
 Db 1238 AGTGTCCGACGACGCTGCTGCTGGCAAGCAATTCAGCTCTCTCTCAAAGCTTTTGTCT 1297
 Qy 303 eLeuIleLeuValLeuGlyValValGlyLeuSerAlaLeuThrTyRMetGlyLeuIleAl 323
 Db 1298 GGCCTGATTCCTGCTCAGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1357
 Qy 323 aProTrpThrGlyArgPheTyRSerLeuTrpAspThrAsnTyRAlaLysIle----- 340
 Db 1358 CCCCCTCGGTGGAAGATTTTATCTCTTTGGGATCTGGCTATGCGAGGTCCACAGTGA 1417
 Qy 341 -----His-IleProIleIleAlas 347
 Db 1418 GTCAAAATGCCATACCTTCGGGATCTATATTTATTTTCAACATAGTCCCATTTATGCT 1477
 Qy 347 eValSerGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeu 367
 Db 1478 CGGTCTCCGAACACACGACCCCGCTTGGCCCTCATTTCTACTTGTACCTCGAATGCTTA 1537
 Qy 367 leTrpLeuPheProAlaGlyValPheLeuLeuPheLeuAspLysAspGluHisValP 387
 Db 1538 TCTTCTTTTCTCCCTCGCGGTCTCTTGTGNGTTCCTCAAGAGAGCTTCGCGATGACAGAT 1597
 Qy 387 heValIleAlaTyRSerValLeuCysSerTyRTrpPheAlaGlyValMetValArgLeuMet 407
 Db 1598 TCATCATCATTTATGCCGTTCTCAGTGCCTATTTCGGGTGTCATGCTTGTGCTTATGC 1657
 Qy 407 eThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspI 427
 Db 1658 TTGTCAATCAGGCTGTGTGTGTGTTCTCTCCCGCATTTGGCTTCTCCAAACTTCTCGAG 1717

1555 TATCTGATCATTCATACATAACATGAGAGCGTTGTAGTTTCCACCACCTGAAATCTCTAC 1614
 159 pAlaserAlaGlyLeuLeuAlaAlaGlyPheLeuAlaValProGlyTyrLeuSerAr 159
 1615 ACCATCAGCTGCTATTTGGCGCGCTTTCATTCGATGATGATGATGATGATGATGATG 1674
 159 gSerValAlaGlySerTyrAspAlaGlyLeuAlaAlaLeuLeuMetValThrPh 179
 1675 ATCTGTGCGCGCTTTCATGACACGAGCAATGCGCATCTTCTCTGATGAGCTCTT 1734
 179 eMerPheThrPheLeuValAlaGlyLeuLeuSerLeuLeuMetHisAlaThrCysAlaAla 199
 1735 CTACTCTGGATTAAGCGCGTCAAAACCGGATGCTCATTTTGGGATGATGATGATGATG 1794
 199 uPheThrPheTyrMetValSerAlaTyrGlyTyrValPheLeuLeuMetValThrAsnLeu 217
 1795 GTTCTACGGTGGATGTTGTCATGGGTGGTGTAGCTTTTATCATCCCAACAG-TATGT 1853
 218 -----IleProLeuHisValPh 223
 1854 CGCTCGGCCCTCAATTCGATGCTGTTTACTCTTTTTCAGTATTCATTCGACGCTT 1913
 223 eLeuLeuLeuLeuMetGlyArgTyrSerSerLeuTyrSerAlaTyrThrTyrPhe 243
 1914 TGTCTCATTTGTCATGGCGAGTTCACAAACCGGCTTTATACCGCTTACTCTCTCTGCTA 1973
 243 rAlaAlaGlyThrValAlaSerMetGlnLeuPheValGlyPheLeuProIleArgSe 263
 1974 TGTCTTGGAACTATCGCTCCATGACGCTGCTTTTGGAGTCTCTCCCTCCATCCGAC 2033
 263 rAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuLeuGlnLeuValAlaPheGlyAs 283
 2034 CTCTGAGCACATGGCGGCTTGGGTGTTTCGGTCTTGTACAGCTGATCGGATCTGCTGA 2093
 283 pPheValGlyGlnLeuSerThrAlaLysPheLysValIleMetMetValSerLeuPh 303
 2094 ACTGTCGACGACTGCTGCTGGCAGCAATTCAGCTCTCTTCAAGACTTTTGTGT 2153
 303 eLeuLeuValLeuGlyValGlyLeuSerAlaLeuThrTyrMetGlyLeuLeuAl 323
 2154 GCGCGTATTCCTGCTAGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2213
 323 aprThrPheGlyArgPheTyrSerLeuTyrAspThrAsnTyrAlaLysIle 340
 2214 CCCCTTCGCTGAGAGATTTATTTCTTTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2273
 341 -----His-IleProIleLeuAlas 347
 2274 GTCAAAATGCCATACCTTCGCGATCTATATTTATTTGCAACATAGTGGCCATTTGCT 2333
 347 erValSerGluHisGlnProValSerThrProAlaPhePheAspThrHisPheLeu 367
 2334 CCGTCTCGAAACACACCCACCTTGGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2393
 367 leThrLeuPheProAlaGlyValPheLeuLeuPheLeuLeuLeuLeuLeuLeuLeuLeu 387
 2394 TCT 2453
 387 heValIleAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValAlaLeuMetL 407
 2454 TCATCATATTATGCGGCTTCTGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2513
 407 euThrLeuThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspI 427
 2514 TTGTCTACGCTGTTGTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2573
 427 leTyrLeuAsp- 430
 2574 CGTATTATGACCCGCTCATCCCCGAAAGCAGACGAGAGCTGGCGAGTCTCAGACGCGAG 2633
 431 -----PheLysThrSerAspArgLysTyrAlaLeuLysProLa 443
 2634 TTGTCTCCAGTCCAAAGCGGAGAGATGGCGCTGCCAAGCCCAATAAGAGCGGCTTCT 2693

444 -----AlaLeuLeuAlaLysLeuLeuValSerGlySerPhe- 455
 2694 CTTTCACAGGTATTTGAGCGCAAGTCTGCTCGGCATCTTTGGTCTCGACATCGAT 2753
 456 -----llePheTyrLeuTyrLeuPheValPheHisSerThrTrp 469
 2754 TTGCTGTGGTTTCCATCTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2813
 469 alThrArgThrAlaTyrSerProSerValValLeuProSerGlnThrProAspGlyL 489
 2814 TGCATTCACAGGTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2873
 489 ysLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 509
 2874 GCCAAATATCATTCATGATTTCCGAGAGGCTTACTTGTGATTCGCAAAACACCGCCG 2933
 509 luAspSerLysValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGlyGlyMetAlaAspA 529
 2934 AGACAGCTCATGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2993
 529 rgThrThrLeuValAspAsnThrTrpAsnAsnThrHisIleAlaLeuValGlyLysA 549
 2994 GCCCACCTCTTGTGATACCAATACCTGGAATAACACCCACATTCGCCAGTTGGTAAG 3053
 549 laMetAlaSerProGluGluLysSerTyrGluLeuLeuLysGluHisAspValAspTyrV 569
 3054 CCATGCTTCCAAAGAGTCTCGATATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3113
 569 alLeuValIlePheGlyLeuLeuGlyPheGlyGlyAspAspIleAsnLysPheLeuT 589
 3114 TTCTTGTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3173
 589 rMerIleArgIleSerGluGlyLeuTrpProGluGluLeuLysGluArgTyrPheTyr 609
 3174 GGATGTTAGATCTTCAAGGTGAATGCTGACGAGGTGACAGAGTGCAGAGTCACTACTT 3233
 609 hrAlaGluGlyLeuTyrArgValAspAlaArgAlaSerGlu- 622
 3234 CTCAAAGAGGGAGTATCTGTCGATGACAGGGGCTG-CGTCTATTTTGTATTATTGTC 3292
 623 -----ThrMetArgAsnSerLeuLeuTyrL 631
 3293 TGGCGAAATTTGCTGACAACTCTGGCAGCACCCCTTCTTCTTCTTCTTCTTCTTCTTCT 3352
 631 ysMetSerTyrLys- 635
 3353 AAATGTTTACTA-CCGCTAGTCAAAATACATTTCTGATAGTGAACATCGTCTGACAT 3411
 636 -----AspPheProGlnLeuPheAsnGlyGlyGlnAlaThrAspArgValArgGlnGlnM 654
 3412 ATCTCAGCTTCTCCGAGCTTATGTTGACACCCCGCTCAAGACAGAGTTTCGAGGCCAAA 3471
 654 etIleThrProLeuAspValProProLeuAsp-TyrPhe- 666
 3472 TTATCCTCTCTAACAGTGT---ACTCTTGTACTCTTGTGTAAGTTTAGGGTTTGAAGAT 3528
 667 -----AspGluValPheThrSerGluAsn 674
 3529 AGTGTATAGACGCTAATGCTGTCACATCTCTGACAGCAGAGGCTTTCATCCGAAAT 3588
 675 TrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGlyArgThrLeuArgAsp 694
 3589 TGGATCGTCAGGATCTCAAGGTCAAGAGGAAGATCCCATTTGACGAGACCAAGGCC 3648
 695 Val- -----GlyGluLeuThrArgSerThrLysThrArgArgSerIle 709
 3649 GTTACTGCTCGAAACGGGGTAAAGAGTTGAAGAGTCTCTAGTCCAGTGGGGCGTG 3708
 710 LysArg 711
 3709 AAGCGG 3714

| | | |
|------|---|------|
| 534 | AspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerPro | 553 |
| 1790 | GATATAATACGTGGAAACATATGCTCATAGCGCTGGTGGCAAGCAATGTCTTCAACC | 1849 |
| 554 | GlulGlyLysSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePhe | 573 |
| 1850 | GAGGAGAGTCTACGAAATTATGACATCTCTGACGTGGACTACGTTTGGTGACTTT | 1909 |
| 574 | GlyClyLeuIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTyrMetIleArgIle | 593 |
| 1910 | GGCGGTGTGATCGGCTATTCTGCGCATGATATCAACAAGTTCCGTGTGATGTCGGAATT | 1969 |
| 594 | SerGluGlyIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGlu | 613 |
| 1970 | GCTCAGGAGAGCATCCCAAGGACATTAAAGAAAGCGATTACTTACGACCGCGGTGAA | 2029 |
| 614 | TyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSer | 633 |
| 2030 | TTCAGGTAGATCCGGAAGTGTCTCGGCGCTGCTCAACTGCCATTATGTACAAATTAAGC | 2089 |
| 634 | TyrLysAspPheProGlnLeu-----PheAsnGlyGlyGlnAlaThrAspArgVal | 650 |
| 2090 | TACTACAGATTCGGGAAATTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGCAC | 2149 |
| 651 | ArgGlnGlnMetIleThrProLeuAspValProProLeuAspTyrPheAspGluValPhe | 670 |
| 2150 | CGTAACCGCGTCAATGGGAATAAGGACTTC---GATCTGAACCTACTGGAGAGGCGCTAC | 2206 |
| 671 | ThrSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAsp----- | 685 |
| 2207 | ACCACAGAACTCGGTGTGTCGCATCTATAGGTGTAAGAGCCGCATGAGTTCAATAGA | 2266 |
| 686 | -----AspAlaGlnGlyArgThrLeuArgAspValGlyGluLeuThrArgSerSer | 702 |
| 2267 | CCATCACTGAAGCAACGAGGAGAACGATCTCTCCAGCAACTTCATTTCGAGAAAGAC | 2326 |
| 703 | ThrLysThrArgArgSerIleLysArg | 711 |
| 2327 | TCTAAGCGTCCGAAAGCGCTACATACGA | 2353 |

RESULT 15
US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, Gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

Page 12

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 558.27 Seconds
(without alignments)
4707.175 Million cell updates/sec

Title: US-10-028-384-6

Perfect score: 3761

Sequence: 1 MGSDRSCVLSVFQILKLV.....TRSTKTRRSIKRPELGLRV 718

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CNG 1 1 13960 @runat_14122004_131519_8030 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_btc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 2013.5 | 53.5 | 2709 | AK018758 | BC044321 Mus muscu |
| 2 | 1998 | 53.1 | 4513 | BC044321 | BC044321 Xenopus l |
| 3 | 1977.5 | 52.6 | 2669 | AK012153 | AK012153 Mus muscu |
| 4 | 1973.9 | 51.6 | 3802 | AK030363 | AK030363 Mus muscu |
| 5 | 1932 | 51.4 | 2730 | AK081547 | AK081547 Mus muscu |
| 6 | 1928.5 | 51.3 | 2656 | BC028897 | BC028897 Mus muscu |
| 7 | 1919.5 | 51.0 | 3643 | AK077877 | AK077877 Mus muscu |
| 8 | 1836 | 48.8 | 2118 | AY418284 | AY418284 Homo sapi |
| 9 | 1812.5 | 48.2 | 2727 | AK087470 | AK087470 Mus muscu |

| | | | | | | |
|----|--------|------|------|---|----------|--------------------|
| 10 | 1807.5 | 48.1 | 2103 | 9 | AY418286 | AY418286 Mus muscu |
| 11 | 1689 | 44.9 | 2419 | 3 | CNS0ADYZ | EX816490 Arabidops |
| 12 | 1465 | 39.0 | 2079 | 9 | AY418285 | AY418285 Pan trogl |
| 13 | 1430.5 | 38.0 | 1088 | 9 | CNS06CLO | AL392680 T7 end of |
| 14 | 1404.5 | 37.3 | 2334 | 3 | AY310157 | AY310157 Rattus no |
| 15 | 1303.5 | 34.7 | 981 | 9 | CNS06YOT | AL420435 T3 end of |
| 16 | 1138 | 30.3 | 972 | 9 | CNS07ABF | AL436261 T7 end of |
| 17 | 1100.5 | 29.3 | 869 | 9 | CNS06DY5 | AL394419 T3 end of |
| 18 | 1009.5 | 26.8 | 978 | 7 | CR413615 | CR413615 |
| 19 | 946.5 | 25.2 | 1141 | 6 | CD505084 | CD505084 CDA71-G02 |
| 20 | 932 | 24.8 | 904 | 7 | CF821733 | CF821733 |
| 21 | 929.5 | 24.7 | 1345 | 7 | CR606556 | CR606556 full-len |
| 22 | 916 | 24.4 | 897 | 3 | CO008636 | CO008636 EST796971 |
| 23 | 911 | 24.2 | 836 | 7 | CO012548 | CO012548 EST800883 |
| 24 | 910.5 | 24.2 | 1019 | 5 | EX404479 | EX404479 EX404479 |
| 25 | 896.5 | 23.8 | 960 | 6 | CD793779 | CD793779 EST665140 |
| 26 | 895 | 23.8 | 852 | 7 | CN458732 | CN458732 UI-M-HB0 |
| 27 | 894.5 | 23.8 | 876 | 6 | CD794446 | CD794446 EST665807 |
| 28 | 894.5 | 23.8 | 878 | 6 | CD788585 | CD788585 EST659946 |
| 29 | 894.5 | 23.8 | 891 | 6 | CD778215 | CD778215 EST649576 |
| 30 | 894.5 | 23.8 | 919 | 7 | CR439276 | CR439276 CR439276 |
| 31 | 884.5 | 23.5 | 960 | 7 | CN062150 | CN062150 Ag2_D35_1 |
| 32 | 862 | 22.9 | 1145 | 5 | EX446691 | EX446691 |
| 33 | 856.5 | 22.8 | 863 | 7 | CR440471 | CR440471 |
| 34 | 852 | 22.7 | 933 | 5 | BUS13073 | BUS13073 AGENCOURT |
| 35 | 852.5 | 22.7 | 862 | 5 | EX713937 | EX713937 EX713937 |
| 36 | 852.5 | 22.7 | 863 | 7 | CR442100 | CR442100 CR442100 |
| 37 | 852 | 22.7 | 1124 | 7 | CK030365 | CK030365 AGENCOURT |
| 38 | 851 | 22.6 | 823 | 7 | CF736482 | CF736482 UI-M-HD0 |
| 39 | 848.5 | 22.6 | 836 | 5 | EX741571 | EX741571 |
| 40 | 845 | 22.5 | 836 | 5 | CNS0172X | CNS0172X |
| 41 | 829.5 | 22.1 | 903 | 6 | CA487501 | CA487501 Drosophl |
| 42 | 826 | 22.0 | 931 | 7 | CO008635 | CO008635 EST796970 |
| 43 | 824 | 21.9 | 593 | 8 | BZ295215 | BZ295215 CG1357.r1 |
| 44 | 821 | 21.8 | 950 | 5 | BUS15954 | BUS15954 AGENCOURT |
| 45 | 813.5 | 21.6 | 894 | 5 | BUS16368 | BUS16368 AGENCOURT |

ALIGNMENTS

RESULT 1

AK018758

LOCUS

DEFINITION

AK018758

2709 bp

mRNA

linear

HTC 03-APR-2004

Mus musculus adult male liver cDNA, RIKEN full-length enriched

library, clone:130006C19 product:hypothetical Oligosaccharyl

transferase (OTase) STT3 subunit containing protein, full insert

sequence.

AK018758

AK018758.2

GI:26384577

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

REFERENCE

1

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalizaton and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

PUBMED

11042159

REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,

Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)

20530913
11076861

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)

6 (bases 1 to 2709)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Noma, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, C., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)

On Dec 10, 2002 this sequence version replaced gi:12858635. Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTNN 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

Location/Qualifiers
1..2709
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:1300006C19"
/db_xref="taxon:10090"
/clones="1300006C19"
/sex="male"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
76..2547

FEATURES
source

CDs

/protein_id="BAB31390.1"
/db_xref="GI:12858636"
/translation="MAEPSAPESKHKSSLNSSPMSGLMALNSRHHGHGPGTQASRA
AAPKPPAGISGGISQPAQWQSLLSFTILFLAWLGSFSLFAVIRPESIIHFDPP
FNVRSTHLAGHGFYFLNWRDRAWYPLGRVTVGLMTAGLTHWLTINT
VHTRDVCVFLAPTFESLTSISTFLATRLMNGAGLAACTAIAPVGVISRSVAGSFD
NGEIALFALQFTYIYLWKSVKTSVFVMTCCCLSYFVMSAWGVVFIINLIPHFV
LLLMYRSKRVYIATSTLYIIGLILSLMQIPFVGPOPIRTSEHMAAGVSLMDTGYAKI
QYRLRLTQKEQFTLFFGLSAGAVLSVLYITGTIAPFWSRFFVSLMDTGYAKI
HPIPIASVSEHQFTTFFSFFDLHLVCTFFAGLMFCIKNIDERSFVSLTASAVIF
AGVRLMLTTPVVMLSAFAFNVFHYLGDMMKRENPVPSDDSDDDKRNPNLY
DKAGVKRHVTQEKEPBGNGIKSIIVTLMMLLMFAVCTVTSNAYSSPSVL
ASVNDGTENIILDPREAVFMRONTDEHARVMSWDYGYOIGANMTTILVDNNTN
NSHIALVXAMSSNETRAYKIMRSLDVLVIFGVVIGYSGDDINKFLMVRVLAEGE
HPRDLREGDYFTQQGSEFVDRKAGSTLNCIMYKSYRREGEMOLDDEPTPGFORTRN
AIGNKDIKFKLEAFTSEHVLRIYIKVAPDNRETLCHEPRTVINIPVKQKLSKKT
TKRKGYVKNKLVFKKKTSTKT"
2693..2698
/note="putative"
2709
/note="putative"

polyA_signal
polyA_site

ORIGIN

Alignment Scores:
Pred. No.: 2,04e-213 Length: 2709
Score: 2013.50 Matches: 399
Percent Similarity: 69.05% Conservative: 112
Best Local Similarity: 53.92% Mismatches: 177
Query Match: 53.54% Indels: 53
DB: 3 Gaps: 12

US-10-028-384-6 (1-718) x AK018758 (1-2709)

Qy 12 PheGlnThrIleuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSer 31
Db 268 TGGCAGCGTGTGCTTCCTTCACCACTCTCTCGCTGGCTGGCGCGCTTACGCTCG 327
Qy 32 ArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsn 51
Db 328 CGCTCTTCGCGCTCATCCGCTTCGAGAGCATCATCCAGAGTTCGACCGCTGGTTTAC 387
Qy 52 TyrArgAlaThrIleLysLeuValAlaAsnSerPheTyrLysPheLeuAsnTrpPheAsp 71
Db 388 TATAGATCAACACATCATCTTCATGAGTTCATGAGTTCATGAGTTCATGAGTTCAT 447
Qy 72 AspArgThrTrpTyrProLeuGlyValThrGlyThrLeuTyrProGlyLeuMet 91
Db 448 GAAAGAGCATGGTACCCACATGGGAGAGATAGTGGTGGCACCCTTTACCCAGGTTGATG 507
Qy 92 ThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIle 111
Db 508 ATAACAGCTGGCTTATT---CATGGATTTTAAATACATGAAACATACAGTTCACATA 564
Qy 112 ArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyr 131
Db 565 AGAGATGGTGTGTATCTCTTGCACCACTTTAGCGGCTTACATCATCATCATCTTCGTT 624
Qy 132 GluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAla 151
Db 625 CTGCTAACTAGAGAACTGTGGAAACCAAGAGCAGGAGTCTCTAGTCTGCTTCATTGCT 684
Qy 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
Db 685 ATCGTACAGGGTACATATCTCGGTCACTGGCGGGATCTCTTGTATATGAAGCATTCGC 744
Qy 172 IleThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIle 191
Db 745 ATTTTGGGCTTCAGTTCACCTTACTTATGGTAAAGTCTGTGAGACCGGCTGTGTG 804
Qy 192 MethHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyGlyTyr 211
Db 805 TTCTGGACATGTCTGCTGCTTGTCTCATATTTCTACATGCTCTGCTGGGAGGGTAT 864

QY 12 PheGlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSer 31
 DB 228 TGGCAGCTCGTTCCTCTCCATCCCTCTCTCCGCTGCTGCGCGCTTCAGCTCG 287
 QY 32 ArgLeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProIlePheAsn 51
 DB 288 CGCCTCTCGCGCTCATCCGCTTCGAGAGCATCATCCAGAGTTCGACCGGTGTTAAC 347
 QY 52 TyrArgAlaThrIleLysLeuValAsnAsnSerPheTyrLysPheLeuAsnThrPheAsp 71
 DB 348 TATAGATCAACACATCATCTTGCATCTCATGGATTCATGAGTTCCTAAATGGTTGAT 407
 QY 72 AspArgThrIlePheProLeuGlyArgValThrGlyLysThrLeuTyrProGlyLeuMet 91
 DB 408 GAAAGACATGTTACCCACTCGGAGAAATAGTGGTGGCCACCGTTTACCCAGGGTTGATG 467
 QY 92 ThrThrSerAlaPheIleThrHisAlaLeuArgAsnThrLeuGlyLeuProIleAspIle 111
 DB 468 ATACACAGCTGGCTTATT---CATGGATTTA---AATACATTGAACATACACAGTTCCACATA 523
 QY 112 ArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaThrAlaThrTyr 131
 DB 524 AGAGATGTGTGTATTCTCTTGCACCAACTTTTAGCGGCTTACATCCATATCTACGTTTC 583
 QY 132 GluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAla 151
 DB 584 CTGCTAACTAGAGAACTGTGGAAACAGGAGCAGGACTTCTAGCTGCTTGCCTTCATTGCT 643
 QY 152 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAla 171
 DB 644 ATCGTACCAGGTCATATCTCGTCACTGGCGGATCTTTGATTAATAGGAGCATTCGCC 703
 QY 172 IleThrLeuLeuMetValThrPheMetPheThrIleLysAlaGlnLysThrGlySerIle 191
 DB 704 ATTTTGGCTTCAGTTCACCTTACTTACTATGGTAAAGTCTGTGAAGCCGGCTCTGTG 763
 QY 192 MetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaThrIleGlyTyr 211
 DB 764 TTCTGGACAATGCTGCTGCTGTTCATATTTCTACATGCTCTCTCGGTGGGAGGTTAT 823
 QY 212 ValPheIleThrAsnIleLeuPheLeuHisValPheLeuLeuIleMetClyArgTyr 231
 DB 824 GTGTTCATCATCAACCTCATCCCTCTCCATGTTTGTGTGCTGTGATGAGAGGTAC 883
 QY 232 SerSerLysLeuTyrSerAlaThrThrIleThrIleAlaIleGlyThrValAlaSerMet 251
 DB 884 AGCAAGAGAGTCTACATAGCATATAGCATTTCTACATGCTGGGTAAATATATCCATG 943
 QY 252 GlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGly 271
 DB 944 CAGATACCTTTTGTGGGATTTTCAGCCAAATCAGCACAGGAGGACCATGTCAGCTCGAGT 1003
 QY 272 ValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerTh 291
 DB 1004 GTCTTTGGCTGCTGCAAGCTTACGCTTTTTCAGATATCTGAGAGACCGGTTCCACAAA 1063
 QY 291 rAlaLysPheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValVa 311
 DB 1064 ACAGGAGTTCAGACCCCTTTCTTTT---TTGGGTGCTCTACGTGTCAGGCGCTGT 1117
 QY 311 lGlyLeuSerAla-----LeuThrTyrMetGlyLeuIleAlaProThrIleGlyArgPh 329
 DB 1118 GTTCTCTAGTCTCATCTCTGACATACACAGGTTTATTTGACCATGAGTGGCGAGGT 1177
 QY 329 eTyrSerLeuThrAspThrAsnThrValAlaLysIleHisIleProIleAlaSerValSe 349
 DB 1178 TTATTCATATGGGATCTGGGTATGCAAAATACATTCATTCATATTCATCATGCTC 1237
 QY 349 rGluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeu 369
 DB 1238 TGAACATCAGCCTACGACATGGGTGTCTTTCTTTCTTTGATCTACATATCTTTGTATGAC 1297

QY 369 UpheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIle 389
 DB 1298 CTTCCAGAGCCCTATGTTCTGCATCAAAAATATCAAGATGAAGAGTATTTGTCCG 1357
 QY 389 eAlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrIle 409
 DB 1358 TCTGTATGCGATCAGTGTCTTACTTTGCGGAGTGTGTGCGGTGATGCTGACTCT 1417
 QY 409 UpheProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrIle 429
 DB 1418 GACCCCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
 QY 429 u-----AspPheLys-----ThrSerAsp----- 435
 DB 1478 GGGGGGTGACATGAAGGAAACCCACCTGTGTGAGGACAGCAGTGTATGAGGATGACAA 1537
 QY 436 -----ArgLysTyrAlaIle----- 440
 DB 1538 AAGAAACCCAGAAACTTGTATGACAGGAGGTAAGTGGAGGATGTGTACAGAGCA 1597
 QY 441 ----LysProAlaAlaLeuLeuAla-----LysLeuIleValSerGlySerPheIle 456
 DB 1598 AGAGAAACCTGAAGAGGCTTGGGCCCCCAACATCAAAAGCATTTGTACCATGCTGATGCT 1657
 QY 456 ePheTyrLeuTyrLeuPheValPheHisSerThrTyrValThrArgThrAlaTyrSerSe 476
 DB 1658 CATGCTCCGATGATGTTTCCGGTCCACTGCACGTGGGTCAAGCAACGCCCTACTCCAG 1717
 QY 476 rProSerValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspPhe 496
 DB 1718 TCCAAAGTGGTCTTGCCTCTCTCAATCATGATGATGATGATGATGATGATGATGAT 1777
 QY 496 eArgGluAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTr 516
 DB 1778 TTAGAGAGCTTACTTTGGCTGACACAAACACCGATGAACACGCCGCTCATGCTGCTG 1837
 QY 516 pTrpAspTyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAs 536
 DB 1838 GTGGGACTACGCTATCAGATTGCTGCGATGCGCCACACAGGACCATCTCTGTGTGATACAA 1897
 QY 536 nThrTrpAsnAsnThrHisIleAlaIleValIleGlyLysAlaMetAlaSerProGluGly 556
 DB 1898 CACCTGGACACACACCCATCGCATGCTGCGAAAGCTATGCTTCCATGAACGGC 1957
 QY 556 sSerTyrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyLe 576
 DB 1958 CGCCTATAAAATCATAGGCTCCCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2017
 QY 576 uIleGlyPheGlyGlyAspAspIleAsnLysPheLeuTyrMetIleArgIleSerGlu 596
 DB 2018 GATTGGCTATTCGGGGACGATATCAACAGTTCTCTGGATGCTGAGGATAGCTGAGG 2077
 QY 596 yIleTrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgVa 616
 DB 2078 GGAGCATCCCAAGACATCCGGAGGTGACTATTTCACCAGCAGCAGGAGGTTCGGAGT 2137
 QY 616 lAspAlaAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAs 636
 DB 2138 AGACAAAGCTGGTCTCTCTACTCTGTAACTGCTTATGTATAAAATGTCATCTACAG 2197
 QY 636 pPheProGlnLeu-----PheAsnGlyGlyGlnAlaThrAspArgValArgGlnI 653
 DB 2198 ATTTGAGAAATGACAGTAGATTTCGCACCTCCCGGCTTTGACCGACACAGTAAATGC 2257
 QY 653 nMetIleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerG 673
 DB 2258 TGAGATTGGAATAAAGACAT---AAATTCACGATTTGGAGGAGCTTTTACATCAGA 2314
 QY 673 uAsnThrMetValArgIleTyrGlnLeuLysGlyAspAlaGlnIleArgThrLeuAr 693
 DB 2315 GCACCTGGCTTGCAGATATATAAAGTGAAGACACCTGACAA-CAGGAGAC----- 2365
 QY 693 GasPValGlyGluLeuThrArgSerSerThrLysThrArgSerIleLysArgProG 713

| | | | | |
|----|------|--|----------|---|
| Db | 2366 | -----ACTAGTCAAACTCGAGTCAACCA-----CATCTGCCCA 2403 | JOURNAL | Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) |
| Qy | 713 | u 713 | COMMENT | CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. |
| Db | 2404 | A 2404 | FEATURES | Location/Qualifiers URL: http://genome.gsc.riken.jp/ URL: http://phantom.gsc.riken.jp/ 1..3802 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM DB:5330402C17" /db_xref="taxon:i0090" /clone="5330402C17" /sex="male" /tissue type="pituitary gland" /clone_lib="RIKEN full-length enriched mouse cDNA library" 133..2250 /note="unnamed protein product; OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus] (SWISSPROT P46978, evidence: FASTA, 99.7%ID, 100%length, match=2115) putative" /codon_start=1 /protein_id="BAC26921.1" /db_xref="GI:26326355" /translation="MTKGLFRLSYERKQDTLLKLLLSMAAVLSFSLFAVLRPESV IHEPDPYNTFRTRFLAEFGYFHNFDRAWYPLIGGIYPLMTIAAIVYH LPHFTIDIRNVCFPLAPFSFTTIVTHTKELKADAGLAAATVAVPVYLSR SVAGSYDNEGIAIFCMLLTVMYKATGTSIYMAAKCALAVYVSSMGVYFGLNL IPIHVLMTGTREHRIYVAYCTVCLGTSLSMOISFVGFQFVLSSEHMAAFVPEGL QTHAFVDYISKLNFCQFELPFRSVISLVGVFLLVGLLMLTKISFWTRFVSLLL DQSHAKNFIILASVSEHQDTWSSYIFDLQLLVFPVGLYICFNSLSDARIFIMY PKNVEFASVAVMLDLVAPVWGLSGIGVSVQVSVYMKLDSIRPKSKKQDQSTY PIKNVEFASVAVMLDLVAPVWGLSGIGVSVQVSVYMKLDSIRPKSKKQDQSTY YMLRHTNPEDAKVMWDYQYITANRITLVDNNTNTHSRVGMASSTBEKA YEIMRELDVSVLVIFGLTGYSDDDINKFLMVRICGSGTETGRHKENDYVTPGEF RVDRSGSPVLLNGLVYKVCYRFGQVYTEAKRPPGDRVRNAEIGNKDFELVLEAY TTEHWLVRIYKVDLNRGLSR" |
| Db | 2404 | A 2404 | ORIGIN | Alignment Scores: Pred. No.: 7,61e-205 Length: 3802 Score: 1939.00 Matches: 378 Percent Similarity: 69.64% Conservative: 106 Best Local Similarity: 54.39% Mismatches: 183 Query Match: 51.56% Indels: 28 Gaps: 10 US-10-028-384-6 (1-718) x AK030363 (1-3802) |
| Db | 2404 | A 2404 | Qy | 14 ThrilleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu 33 178 ACACCTTCAAGCTTCATCTCGATCGCTGCTGTATCTTTTCTACTCGCTT 237 |
| Db | 2404 | A 2404 | Qy | 34 PheAlaValIleLysPheGluSerIleIleHieGluPheAspProTrpPheAsnTrpArg 53 238 TTTCGTGCTGAGATTGAAGTGCATCCATCCATGAGTTTATTAATTCATTAATTCGG 297 |
| Db | 2404 | A 2404 | Qy | 54 AlathrlYstYrLeuValAsnAenSerPheTyIysPheLeuAsnTrpPheAspArg 73 298 ACTACCGGTTTCGTGCTGAGAGGGGTTTATAAATTCATTAATTCATTAATTCGG 357 |
| Db | 2366 | -----ACTAGTCAAACTCGAGTCAACCA-----CATCTGCCCA 2403 | JOURNAL | Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) |
| Qy | 713 | u 713 | COMMENT | CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. |
| Db | 2404 | A 2404 | FEATURES | Location/Qualifiers URL: http://genome.gsc.riken.jp/ URL: http://phantom.gsc.riken.jp/ 1..3802 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM DB:5330402C17" /db_xref="taxon:i0090" /clone="5330402C17" /sex="male" /tissue type="pituitary gland" /clone_lib="RIKEN full-length enriched mouse cDNA library" 133..2250 /note="unnamed protein product; OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus] (SWISSPROT P46978, evidence: FASTA, 99.7%ID, 100%length, match=2115) putative" /codon_start=1 /protein_id="BAC26921.1" /db_xref="GI:26326355" /translation="MTKGLFRLSYERKQDTLLKLLLSMAAVLSFSLFAVLRPESV IHEPDPYNTFRTRFLAEFGYFHNFDRAWYPLIGGIYPLMTIAAIVYH LPHFTIDIRNVCFPLAPFSFTTIVTHTKELKADAGLAAATVAVPVYLSR SVAGSYDNEGIAIFCMLLTVMYKATGTSIYMAAKCALAVYVSSMGVYFGLNL IPIHVLMTGTREHRIYVAYCTVCLGTSLSMOISFVGFQFVLSSEHMAAFVPEGL QTHAFVDYISKLNFCQFELPFRSVISLVGVFLLVGLLMLTKISFWTRFVSLLL DQSHAKNFIILASVSEHQDTWSSYIFDLQLLVFPVGLYICFNSLSDARIFIMY PKNVEFASVAVMLDLVAPVWGLSGIGVSVQVSVYMKLDSIRPKSKKQDQSTY PIKNVEFASVAVMLDLVAPVWGLSGIGVSVQVSVYMKLDSIRPKSKKQDQSTY YMLRHTNPEDAKVMWDYQYITANRITLVDNNTNTHSRVGMASSTBEKA YEIMRELDVSVLVIFGLTGYSDDDINKFLMVRICGSGTETGRHKENDYVTPGEF RVDRSGSPVLLNGLVYKVCYRFGQVYTEAKRPPGDRVRNAEIGNKDFELVLEAY TTEHWLVRIYKVDLNRGLSR" |
| Db | 2404 | A 2404 | ORIGIN | Alignment Scores: Pred. No.: 7,61e-205 Length: 3802 Score: 1939.00 Matches: 378 Percent Similarity: 69.64% Conservative: 106 Best Local Similarity: 54.39% Mismatches: 183 Query Match: 51.56% Indels: 28 Gaps: 10 US-10-028-384-6 (1-718) x AK030363 (1-3802) |
| Db | 2404 | A 2404 | Qy | 14 ThrilleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu 33 178 ACACCTTCAAGCTTCATCTCGATCGCTGCTGTATCTTTTCTACTCGCTT 237 |
| Db | 2404 | A 2404 | Qy | 34 PheAlaValIleLysPheGluSerIleIleHieGluPheAspProTrpPheAsnTrpArg 53 238 TTTCGTGCTGAGATTGAAGTGCATCCATCCATGAGTTTATTAATTCATTAATTCGG 297 |
| Db | 2404 | A 2404 | Qy | 54 AlathrlYstYrLeuValAsnAenSerPheTyIysPheLeuAsnTrpPheAspArg 73 298 ACTACCGGTTTCGTGCTGAGAGGGGTTTATAAATTCATTAATTCATTAATTCGG 357 |

QY 74 ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThr 93
 DB 358 GCTGGTACCTTTGGCCCAATCATTTGGAGGAACAATTTACCCAGGTTTATATGATCACT 417
 QY 94 SerAlaPheIleThrHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsn 113
 DB 418 TCTGCTGCATCTACCATGTACTC---CATTTCTCCATATCACCATTTGACATCCGGAT 474
 QY 114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe 133
 DB 475 GTCTGTGTCTTCTGGCCCGCTCTCTCTCTTTCCACCACCATCGTTACGTACCACCTT 534
 QY 134 ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
 DB 535 ACCAAGAGCTCAAGATGAGGAGCTGGGCTTTCTGCTGCCATGATGCTGTAGTT 594
 QY 154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIleThr 173
 DB 595 CTTGGGTATATTCTCGATCTGTAGCTGGCTCTATGATAAATGAAGGAATGCTATCTTT 654
 QY 174 LeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHis 193
 DB 655 TGCATGCTGCTTACTACTACATCTGCATCAAGCCAGTGAAGACTGGTTCCATCTATTGG 714
 QY 194 AlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPhe 213
 DB 715 GCTGCCAAGTGTGCCCTCGCTTATTCTACATGCTCTCTTCATGGGAGGCTATGTTC 774
 QY 214 IleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSerSer 233
 DB 775 CTGATCACTGATCTCTACATGCTGCTGCTAATGCTGACAGGCGGTTTCTCAC 834
 QY 234 LysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIle 253
 DB 835 CGGATCTACGTAGCTACTGCTACTGTTTACTGCTGGCACCATTCTCTTCTATGAGATT 894
 QY 254 ProPheValGlyPheLeuProIleArgSerAspHisMetAlaAlaLeuGlyValPhe 273
 DB 895 TCTTTGTTGGTTTCCAGCCCGCTTCTCATCAGAACATGCGACGCTTTGGAGTGTT 954
 QY 274 GlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLys 293
 DB 955 GGTCTCTGTACATCCATGCTTTCTGTAGATTACCTGTGCAGCAAGTTGAATCCACAGCAA 1014
 QY 294 PheLysValIleMet-----MetValSerLeu-----PheLeuIleLeuValLeuGly 309
 DB 1015 TTCGAAGTTCTTTTCCGAGGTGTTATCTCCCTGTTGGCTTGTCTCTCTCTCTCTCTCT 1074
 QY 310 ValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPhe 329
 DB 1075 GCTCTC-----CTCATGCTAACAGGAAAAATTTCTCCCTGGACAGGGCGTTTC 1122
 QY 330 TyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSer 349
 DB 1123 TACTCTCTGCTGGATCCCTCTTATGTCTAAGAAATCAATCCCATTTATTCATCTGTCTCT 1182
 QY 350 GluHisGlnProValSerTrpProAlaPhePhePheAspThrHisPheLeuIleTrpLeu 369
 DB 1183 GAGCACAGCCCAACACCTGCTCTCTCTATTTTGTATCTACAGCTCTCTCTCTCTCTCT 1242
 QY 370 PheProAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIle 389
 DB 1243 TTTCCAGTTGGCTCTATTACTGCTTTTAGCAACCTGTCTGTATGCTCGGATTTTATCATC 1302
 QY 390 AlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 409
 DB 1303 ATGTATGTGTGACCAAGCATCTTTCTAGCTGTATGTGGCTGTATGTCTGTATGTATG 1362
 QY 410 ThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyr--- 428
 DB 1363 GCACCTGTATGTGCAVTTCTTCGGCATTTGTGTTTCCAGGTGTGTGCCACATATATG 1422
 QY 429 -----LeuAspPheLysThrSerAspArgLys-----Tyr 438

DB 1423 AAAAATCTGGACATAAGTCGCCCCAGACAGAAAGAGCAAGCAAGCAAGGATTTCTATTAC 1482
 QY 439 AlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyr 458
 DB 1483 CCTATTAGATAGTGGTGGCGAGTGGGATGATACTACTGTC-----ATGGCTTTCTTT 1533
 QY 459 LeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerSerProSer 478
 DB 1534 CTCATCACCTACACCTTTTCTGCTTGGTGGTACAGTGAAGCTTATTTTCTCTCTCC 1593
 QY 479 ValValLeuProSerGlnTrpProAspGlyLysLeuAlaLeuIleAspAspPheArgGlu 498
 DB 1594 ATTGTTCTGCTGCTGCTGGTGGGATGCGCAGTAGGATCATTTTGTATGACTTCCGAGAA 1653
 QY 499 AlaTyrTyrTrpLeuArgMetAsnSerAspGluSerLysValAlaAlaTrpTrpAsp 518
 DB 1654 GGTATTTATTTGCTCCGTCACATATCTCCAGAGGATGCAAAAGTCATGCTTGTGGAT 1713
 QY 519 TyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrp 538
 DB 1714 TATGGCTACCAATTTACTGCAATGCAATCGCAATTTTGTAGTGGACAATAACATGCG 1773
 QY 539 AsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyr 558
 DB 1774 AATAATACCATTTCTCGAGTAGGCGAGCAATGGCATCCACAGAAAGAAAGCCAT 1833
 QY 559 GluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGly 578
 DB 1834 GAAATCATGAGGAGCTTGTATGTCAGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1893
 QY 579 PheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIle--- 597
 DB 1894 TATTTCTCGATGATATCAACAAGTTCTTTGGATGTCGCGATTTGAGGAAGCAACAGAG 1953
 QY 598 TrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAsp 617
 DB 1954 ACAGGAAGACACATTAAGAGAGATGACTACTATCTACTCTGAGGAAATTCGCTGTGAT 2013
 QY 618 AlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPhe 637
 DB 2014 CGTGAGGTTTCTCCGCTGCTCTCACTGCTTATGTACAAATGTCTTACTACCGCTTT 2073
 QY 638 ProGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnGlnMet 654
 DB 2074 GGGCAGGCTCTACAGAAAGCCAAAGCGCTCCACAGGCTTTGACCGTGTGCAAAATGCTGAG 2133
 QY 655 IleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAsn 674
 DB 2134 ATTGGTAATAAGATTTT---GAGCTTGATGCTCTGGAGGAGGCTATACACAGAACAC 2190
 QY 675 TrpMetValArgIleTyrGlnLeuLysLysAspAspAlaGlnGly 689
 DB 2191 TGGCTAGTCAGATATACAGGTAAGGACCTGGATTAATCGAGC 2235
 RESULT 5
 AK081547 2730 bp mRNA linear HTC 03-APR-2004
 LOCUS Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
 DEFINITION library, clone: C130038C18 product: OIGSACCHARYL TRANSFERASE S1T3
 SUBUNIT HOMOLOG (BS) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus
 musculus], full insert sequence.
 AK081547
 ACCESSION AK081547.1 GI:26100006
 VERSION
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1
 REFERENCE Carninci, P. and Hayashizaki, Y.
 AUTHORS High-efficiency full-length cDNA cloning
 TITLE Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL

```

99279253
MEDLINE
PUBMED
REFERENCE
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalisation and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11042159
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
MEDLINE
PUBMED
REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2730)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koyas, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
FEATURES
source
1..2730
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C130038C18"
/db_xref="taxon:10090"
/clone="C130038C18"
/tissue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

```

```

/dev_stage="16 days embryo"
125..2243
/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
(B5) (INTRAGOL MEMBRANE PROTEIN 1) homolog (Mus musculus)
(SWISSPROT)P46978, evidence: FASTY, 99.7%ID, 100%length,
match=2115)
putative"
2717..2722
/note="putative"
2730
/note="putative"
ORIGIN
Alignment Scores:
Pred. No.: 2,75e-204 Length: 2730
Score: 1932.00 Matches: 378
Percent Similarity: 69.68% Conservative: 107
Best Local Similarity: 51.31% Mismatches: 182
Query Match: 51.37% Indels: 29
DB: 3 Gaps: 10
US-10-028-384-6 (1-718) x AK081547 (1-2730)
QY 14 ThrileLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArgLeu 33
DB 170 ACACCTTCTAAAGCTTCTCATCTCGATCGCTGCTGTGTATCTTTTCTACTCTCTT 229
QY 34 PheAlaValIleLysPheGluSerIleIleHisGluPheAspProtrpPheAsnTyArg 53
DB 230 TTTCGTGTGTCAGATTGAAAGTGTCATCCATGAGTTTCATCGTACTTAAATATCGG 289
QY 54 AlaThrLysTyLeuValAsnSerPheTyLysPheLeuAsnTrpPheAspArg 73
DB 290 ACTACCCGGTTCTGGCTGAGGAGGGGTTTATAAATTCATTAACCTGGTTGATGACCGG 349
QY 74 ThrTptTyProLeuGlyArgValThrGlyGlyThrLeuTyProGlyLeuMetThr 93
DB 350 GCTTGTTACCTTGGGCGGATCATTTGGAGGAACATTTTACCAGGTTTAAATGATCACT 409
QY 94 SerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProleAspIleArgAsn 113
DB 410 TCTGCTGCAATCTACCAGTACTC---CATTTCTTCATATCACCATTGACATTCGGAAT 466
QY 114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyGluPhe 133
DB 467 GTCGTGTTTCTCTGGCCGCTTTCTCTCTTACCACCATCGTACCTACCATCTT 526
QY 134 ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
DB 527 ACCAAGAGCTCAAGGATGAGGAGCTGGGCTTCTGCTGCCATGATGCTGATG 586
QY 154 ProGlyTyTrileSerArgSerValAlaGlySerTyAspAsnGluAlaIleThr 173
DB 587 CTTGGGTATATTTCTGATCTGAGTGGCTCTCTATGATTAATGAAGAAATGCTATCTTT 646
QY 174 LeuLeuMetValThrPheMetPheTrpIleValGlnLysThrGlySerIleMetHis 193
DB 647 TGCATGCTGCTTACTTACTATGATGATCAGGAGTGAAGTGAAGTGAAGTGAAGT 706
QY 194 AlaThrCysAlaAlaLeuPheTyTrpPheTyMetValSerAlaTrpGlyTyValPhe 213
DB 707 GCTGCCAAGTGTGCTGCTCTTATTTCTATGATGCTCTTCTTATGAGGAGGCTATGTTTC 766
QY 214 IleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTySerSer 233
DB 767 CTGATCAACTTGAATTCCTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
QY 234 LysLeuTySerAlaTyThrThrTyAlaIleGlyThrValAlaSerMetGlnIle 253
DB 827 CGGATCTACGTAGCTTACTGTACTGTTTACTGCTGGGACCATCTTCTTCTGACAGAT 866
QY 254 PropheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPhe 273

```


74 ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThr 93
318 GCTTGGTACCTTTGGGCGCAATCATTTGGAGGAAACAATTTACCCAGGTTTAATGATCACT 377
94 SerAlaPheLeuTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsn 113
378 TCTGCTGCATCTACCATGTACTC---CATTTCTTCATATCACCATTGACATTCGGAAAT 434
114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe 133
435 GTCGTGTGTTTCTCGCCCGCGCTTTCTCTCTCTTCCACCACCATCGTTAGTACCACTT 494
134 ThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIleVal 153
495 ACCAAGAGCTCAAGAGTGCAGGAGCTGGGCTTCTGCTGCTGCATGATTCGTAGATT 554
154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaThr 173
555 CTTGGGTATATTCTCGATCTGTAGCTGCTCTATGATGATATGAAGAAATTCGTATCTTT 614
174 LeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMetHis 193
615 TGCATGCTGCTTACTACTACATGTGGATCAAGGAGTGAAGATGGTTCATCTATTGG 674
194 AlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPhe 213
675 GCTGCCAAGTGTGCCCTCGCTTATTCTACATGGTCTCTTCATGGGAGGCTATGTGTTT 734
214 IleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGlyArgTyrSerSer 233
735 CTGATCACTGATTCCTCTACATGCTCTGCTGTGTGATGCTGACAGCCGTTTCTCAC 794
234 LysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIle 253
795 CGGATCTAGTACCTACTGCTTACTTACTCGCTGGGACCACTTCTTCTATGACAGATT 854
254 ProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPhe 273
855 TCCTTTGTGGTTTCAGCCGCTCTTCATCAGAACACATGGCAGGCTTTGGAGTGT 914
274 GlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAlaLys 293
915 GGTCTCTGTGCATCCATGCTTCTGTAGATTACCTGGCAGCAAGTTGAATCCACAGCAA 974
294 PheLysValIleMet-----MetValSerLeu-----PheLeuIleLeuValLeuGly 309
975 TTCGAAGTCTTTCCGAGGTGTTATCTCCCTGGTGGCTTTGCTTCCTCCTCAGTGGGA 1034
310 ValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPhe 329
1035 GCTCTC-----CTCATGTCTAAACAGGAAAAATTTCTCCCTGGACAGGCGGTTTC 1082
330 TyrSerLeuTrpAspThrAsnTyrAlaLysIleHisIleProIleIleAlaSerValSer 349
1083 TACTCTGCTGCTGATCCCTCTTATGCTAAGTAATACATTCCTCATATGATGATCTGTTCT 1142
350 GluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeu 369
1143 GAGCACCACCCACCAACCTGCTCTCTCTACTATTGATCTACAGCTCCTTGTCTCATG 1202
370 PheProAlaGlyValPheLeuLeuPheLeuAspLysAspGluHisValPheValIle 389
1203 TTTCAGTTGGCTCTATATCTGCTTTAGCAACCTGCTGATGCTCGGATTTTATCATC 1262
390 AlaTyrSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 409
1263 ATGATGTTGTTGACCAAGATGATCTTTTTCAGCTGTGAATGCTGCTCTAATGCTGTTATG 1322
410 ThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyr--- 428
1323 GCACCTGTTATGTGCTTCTTCTGCAATGGTGTTCCTCCAGGTGCTGCTCACATATATG 1382

429 -----LeuAspPheLysThrSerAspArgLys-----Tyr 438
1383 AAAAATCTGGACATAAGTCCGCCAGACACAGAGCAAGCAAGCAAGGATTCTACTTAC 1442
439 AlaIleLysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyr 458
1443 CCTATTAGATGAGTGGCGAGTGGGATGATACTGGTC-----ATGGCTTCTTT 1493
459 LeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerProSer 478
1494 CTCATACCTACAGCTTTTCTTCTGACTGGGTGACCAAGTGAAGCTTATTCTTCTCCCTCC 1553
479 ValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuIleAspAspPheArgGlu 498
1554 ATTGTTCTGCTGCTGCTGGTGGGATGGCAGTAGGATCATTTTGTATGATTCGAGAA 1613
499 AlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAsp 518
1614 CGGTATTATTGGCTCCGTCAATACTTCCAGAGGATGCAAAAGTCATGTCATGGTGGAT 1673
519 TyrGlyTyrGlnIleGlyMetAlaAspArgThrThrLeuValAlaAspAsnAsnThrTrp 538
1674 TATGGCTACCAATATTACTGCAATGGCAATCGAA-ATTTAGTGACACATTAACATGG 1732
539 AsnAsnThrHisIleAlaIleValGlyLysAlaValAlaSerProGluLysSerTyr 558
1733 AATAATACCCATATTTCTCGAGTAGGCGAGCAATGGCATCCACAGAAGAAAGGCTAT 1792
559 GluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleGly 578
1793 GAATCATGAGGAGCTTGTATGTCAGCTAATGCTGTTGTCATTTTGGAGGCTTACTGG 1852
579 PheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIle--- 597
1853 TATTCTTGGATGATATCAACAAGTTTCTTTGGATGCTCGGATTCGAGGAACACAGAG 1912
598 TrpProGluGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgValAsp 617
1913 ACAGAAAGACACATTAGGAGATGACTACTACTACTCTACTGGGAAATTCGTTGTAT 1972
618 AlaArgAlaSerGluThrMetArgAsnSerLeuTyrLysMetSerTyrLysAspPhe 637
1973 CGTGAGGCTTCTCCGGTGTCTCACTCACTGCTTATGTACAAAATGTGTACTACCGCTT 2032
638 ProGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnGlnMet 654
2033 GGGCAGGTCTACACAGAGCAAGCGCTCCACAGGCTTTGACCGTCTTCGAATGCTGAG 2092
655 IleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluAsn 674
2093 ATTGTTAATAAGACTTT---GAGCTTGTATGCTCTGGAGAGCGTATATACCAAGAACAC 2149
675 TrpMetValArgIleTyrGlnLeuLysLysAspAlaGlnGly-----ArgThr-Le 692
2150 TGGCTAGTCAGATATACAGGTAAAGACCTGATAATCGAGCTTGTCAAGACATAA 2209
692 uArgAspValGlyGluLeuThr-----ArgSerSerThrLysThrArgArgSerIle 709
2210 ACGTACATTTGGTCCCTGAGCATTTATGCTTCGACTGAGCGGCTCATGTTGAGGACGT 2269
709 eLys 710
2270 GAAG 2273
RESULT 8
AY418284
LOCUS
DEFINITION Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
2118 bp DNA linear GSS 17-DEC-2003
ACCESSION AY418284
VERSION AY418284.1 GI:39774244
KEYWORDS GSS.
SOURCE Homo sapiens (human)

Qy 34 PheAlaValIleLysPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 53
Db 213 TTTGCTGCTGAGATTGAAGTGTCTCATCTGAGTTTGATCTGCTATTTATATCGG 272
Qy 54 AlaThrIleTyrLeuValAsnAsnSerPheTyrIlePheLeuAsnTrpPheAspArg 73
Db 273 ACTACCCGGTTTCTGGCTGAGGAGGGGTTTATAAATTCATTAATCTGATGACCGG 332
Qy 74 ThrTrpTyrProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThrThr 93
Db 333 GCTTGGTACCTTTGGGCGGATGATGGAGGACAAATTTACCCAGGTTTATGATCACT 392
Qy 94 SerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArgAsn 113
Db 393 TCTGTGCAATCTACCATGACTC---CATTTCTTCATATATATACCATTCGACATTCGGAAT 449
Qy 114 ValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGluPhe 133
Db 450 GTCTGTGTTTCTGGGCGCGCTTTCTCTCTTTTACCACCATCTGTTAGTACACCTT 509
Qy 134 ThrIleGluIleLysAspAlaSerAlaGlyLeuLeuAlaGlyPheIleAlaVal 153
Db 510 ACCAAAGAGCTCAAGGATGCGAGGAGCTGGGGTTCTTGGTGGTGCATGATTCGTGATT 569
Qy 154 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIleThr 173
Db 570 CTTGGGATATTTCTGATCTGAGCTGGCTCTATGATTAATGAAGGATTCGTATCTTT 629
Qy 174 LeuLeuMetValThrPheMetPheTrpIleGlyAlaGlnLysThrGlySerIleMethis 193
Db 630 TGCATGCTGCTTACTTACTCATGTGATCAAGGAGTGAAGACTGGTTCATCTATTGG 689
Qy 194 AlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyTyrValPhe 213
Db 690 GCTGCCAGTGTGGCTCGCTTATTTCTACATGCTCTTCATGGGAGGCTATGTTTC 749
Qy 214 IleThrAsnLeuIleProLeuHisValPheLeuLeuLeuMetGlyArgTyrSer 233
Db 750 CTGATCACTGATGATTCCTACATGCTCTGCTGCTAATGCTGACAGCGGTTTCTCAC 809
Qy 234 LysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGlnIle 253
Db 810 CGGATCACTAGCTAGCTACTGCTTACTGCTGGGACCATCTTCTATGACAGATT 869
Qy 254 ProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPhe 273
Db 870 TCCCTTTGTGTTTCCAGCGCGTCTTTCATCAGAACACATGCGACGCTTGGAGTGT 929
Qy 274 GlyLeuIleGlnIleValAlaPheGlyAspPheValIleGlyGlnIleSerThrAla-Ly 293
Db 930 GGTCTCTCTCATCTCATCTGCTTTGCTAGATTACTGGGACACACAGTTGAATCCACACA 989
Qy 293 sPheLysValIleMet-----MetValSerLeuPheLeuIleValLeuGlyValVa 311
Db 990 ATTCCGAAGTTTTCGCGAGTGTATCTCCCTGGTGTGCTTCTCCCGCCACACTGTGGG 1049
Qy 311 LGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTrpThrGlyArgPheTyr-Se 331
Db 1050 ACCCTCCCGCTGCTACAC-----GGAAAAATTTCTCCCTGGACAGCGGCTTTTACTC 1103
Qy 331 rLeuTrpAspThrAsnTyrAlaIleHisIleProIleIleAlaSerValSerGluHis 351
Db 1104 TCTGCTGAATCCCTCTAATGCTAAAAAACAATCCCATTTATGCACTGCTTTCTGAGCA 1163
Qy 351 sGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeuPhePr 371
Db 1164 CCAGCCCAACACTGGTCTTCTTACTATTGATCTACAGCTCTGCTTCTTCTGATTTCC 1223
Qy 371 cAlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTy 391
Db 1224 AGTTGCCCTCTATTACTGCTTTTACCAACCTGCTGATGCTCGGATTTTATCATCATGTA 1283
Qy 391 rSerValLeuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPr 411

Db 1284 TGGTGTGACCGACGATGCTACTTTTTCAGCTGTAATGGTGGTCTAATGCTGCTGATTTGGCAC 1343
Qy 411 oValIleCysValSerAla-AlaValAlaLeuSer-----L 423
Db 1344 TGTATGTGCATCTCTTTCTGGGCATTTGGTGTTCCTCCAGGGTGGCTGTCACATATATGA 1403
Qy 423 ysIlePheAspIleTyrLeuAspPheLysThrSer-----AspAtgLyTy 438
Db 1404 AAAATCTGGACATAAGTCGCCCCACACAAGAGAGCAAGAACCAAGGATTTCTACTTAC 1463
Qy 439 AlaIleCysProAlaAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyr 458
Db 1464 CCTATTAGAATGAAGTGGCGAGTGGGATGATGATGTC-----ATGGCTTTCTTT 1514
Qy 459 LeuTyrLeuPheValPheHisSerThrTrpValThrArgThrAla-TyrSerSerProSe 478
Db 1515 CTATCACCCTACACGCTTTTCATTCGACTTGGTGACCGAGTGAAGCCCTATTTCTCTCCCTC 1574
Qy 478 rValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuLeuAspPheArgG 498
Db 1575 CATTTGCTGCTGCTGCTGGTGGGATGGCAGTAGGATCATTTTGTGATGATTTCCGAGA 1634
Qy 498 uAlaTyrTyrTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAs 518
Db 1635 AGCGTATTATGGCTCCGCTCACATACTCCAGAGGATGCAAAAGCTCATGCTGTGGGA 1694
Qy 518 pTyrGlyTyrGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTr 538
Db 1695 TTATGCTCTACCAAAATTTACTGCAATGGCAATCGGACAAATTTTAGTGACAAATACCATG 1754
Qy 538 pAsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGlu-GluLysSer 558
Db 1755 GAATAATACCATATTTCTCGAGTAGGCGCAATGSCATCCACAGAAAGAAAGCCCT 1814
Qy 558 yTrGluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyGlyLeuIleG 578
Db 1815 ATGAATATCAGGAGGCTTGATGTCAGCTATGCTGTGTCATTTTGGAGGCTTACTG 1874
Qy 578 LypheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIleSerGluGlyIle- 597
Db 1875 GGTATTTCTCGGATGATATCAACAGTTTCTTGGATGGTCCGATTTGGAGGAGCACAG 1934
Qy 598 --TrpProGluIleLysGluArgTyrPheTyrThrAlaGluGlyGluTyrArgVala 617
Db 1935 AGACAGGAAGACACATTAAAGAGAAATGACTACTACTACTCTGCTGCGGAATTCGCTGT 1994
Qy 617 sAlaAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAsp 637
Db 1995 ATCGTGAGGGTTCTCCGGTGTGCTCAACTGCCTTATGTACAAATGTGTACTACCGCT 2054
Qy 637 heProGlnLeuPheAsnGlyGlyGln-----AlaThrAspArgValArgGlnIleM 654
Db 2055 TTGGGAGGCTCTACACAGAACCGCTCCACAGGCTTTGACCGGTTTTCGAAATGCTG 2114
Qy 654 etIleThrProLeuAspValProProLeuAspTyrPheAspGluValPheThrSerGluA 674
Db 2115 AGATTGGTAAATAAGACTTT--GAGCTTGTATGCTCTGGAGGAAGCTATATACACAGAAC 2171
Qy 674 snTrpMetValArgIleTyrGlnLeuLysLysAspAspAlaGlnGly 689
Db 2172 ACTGGCTAGTCAGGATATACAGGTAAAGGACCTGGATATATCAGGG 2218

RESULT 10

AY418286 2103 bp DNA linear GSS 17-DEC-2003
LOCUS
DEFINITION Mus musculus ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418286
VERSION AY418286.1 GI:39774246
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus


```

|||||.....|
1677 GAAGAAAAGCTATGAATCATGAGGAGCTTGATGTCAGCTATGCTGTCATTTT 1736
QY GlyGlyLeuIleGlyPheGlyGlyAspAlaAsnLysPheLeuTrpMetIleArgIle 593
Db |||||.....|
1737 GGAGGCTTACTGGGTATCTCGATGATCAACAGTTTCTTGGATGTCGGATT 1796
QY SerGluGlyIle---TrpProGluGluIleGlyGluArgTyrPheTyrThrAlaGluGly 612
Db |||||.....|
1797 GGAGGAAGCACAGACAGGAGACACATTAAGAGAGATGACACTATATCTCTACTCGG 1856
QY GluTyrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMet 632
Db |||||.....|
1857 GAATTCGGTGTGATCGTGAGGGTTCCTCGGTCTCTCACTCCCTTATGTACAAATG 1916
QY SerTyrLysAspPheProGlnLeuPhe---AsnGlyGly-GlnAlaThrAspArgValArg 651
Db |||||.....|
1917 TGTACTACCCCTTGGGCGAGTCTACACAGAGCGTCCACAGCGCTTGGACCGTTCG 1976
QY gGlnGlnMetIleThrProLeuAspValProPheAspTyrPheAspGluValPheTh 671
Db |||||.....|
1977 AAATGCTGAGATTGGTAATAAGACTTT---GAGCTTGATGCTCTCGAGGAGCGTATAC 2033
QY rSerGluAsnTrpMetValArgIleTyrGlnLeuLysLysAspAspAlaGlnGly 689
Db |||||.....|
2034 CACAGAACCTGGCTAGTCAGGATATACAAAGGTAAGGACCTGGATATCGAGGC 2088

RESULT 11
CNSOADYZ 2419 bp mRNA linear HTC 06-FEB-2004
LOCUS Arabidopsis thaliana Full-length cDNA Complete sequence from clone
DEFINITION GSITPGH512C09 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
ACCESSION BX816490
VERSION BX816490.1 GI:42474670
KEYWORDS HTC; GSLT cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 2419)
Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
Unpublished
2 (bases 1 to 2419)
Genoscope.
Direct Submission
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
length
http://www.genoscope.cns.fr/cgi-bin/ggb?source=Arabidopsis.
Location/Qualifiers
1. 2419
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
FEATURES
source

```

```

/db_xref="taxon:3702"
/clone="GSITPGH512C09"
/tissue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT_6"
1. 2419
/gene="Atlg34130"

ORIGIN
Alignment Scores:
Pred. No.: 3,67e-177 Length: 2419
Score: 1689.00 Matches: 337
Percent Similarity: 65.14% Conservative: 119
Best Local Similarity: 48.14% Mismatches: 206
Query Match: 44.91% Indels: 38
DB: 3 Gaps: 11

US-10-028-384-6 (1-718) x CNSOADYZ (1-2419)
QY 13 GlnThrIleLeuLysLeuValIlePheValAlaIlePheGlyAlaAlaIleSerSerArg 32
Db 222 GAGCTCTTACTCCGTATCTCGATCCTTGGTCTCGTCTACATCTTAGCTTTTCATAGCTCGT 281
QY 33 LeuPheAlaValIleLysPheGluSerIleIleHisGluPheAspProTyrPheAsnTyr 52
Db 282 CTCTTCAGTGTCTCCGCTACGATCAATGATCAAGATTCGATTCGATTCGATTCACATC 341
QY 53 ArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrpPheAspAsp 72
Db 342 CGTACGACCTTTCTCTCTCACAGAGAAAGCTTCTACGAGTTTGGAACTGGTTCGATTC 401
QY 73 ArgThrTyrTrpProLeuGlyArgValThrGlyGlyThrLeuTyrProGlyLeuMetThr 92
Db 402 GAGAGTTGTAACCTTAGTTCGATCATCGCGGTACTCTTTATCCGGTCTTATCGGTTC 461
QY 93 ThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIleAspIleArg 112
Db 462 ACAGCTGCGCTTAATCTACTGACATTCGGA---TTTCTCCGATTTCTTCGTTTCATATCCGT 518
QY 113 AsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAlaThrTyrGlu 132
Db 519 GAGTTTGTTGTTTAACCGCACCGTTTTCGCGTCGATACGACTCTCGTTCGATTTTC 578
QY 133 PheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPheIleAlaIle 152
Db 579 TTTGGTAAAGAGCTTTGGGATACAGAGCTGGTCTTGTGTCTGTCTTATCGCTTATTCGCTATT 638
QY 153 ValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAlaIleAlaIle 172
Db 639 TGTCTCGGATGATTTCTGTTCTAGTCTGGTTCGATGATGATACGAGCTGTTCGATT 698
QY 173 ThrLeuLeuMetValThrPheMetPheTrpIleLysAlaGlnLysThrGlySerIleMet 192
Db 699 TTGCTCTCTGCTGACGTTTATCTTCTTGTAAAGCGGTGAACACTGGTTCATTAGCT 758
QY 193 HisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGlyGlyTyrVal 212
Db 759 TGGCTCTGCTCTGCGTTGTTGTTACTTTATATGTTTCAGCTTCGGGAGGATATGTG 818
QY 213 PheIleThrAsnLeuIleProLeuHisValPheLeuLeuLeuLeuMetGlyArgTyrSer 232
Db 819 TTTATTATCAATCTGGTCTCTCTTTACGTGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGT 878
QY 233 SerLysLeuTyrSerAlaTyrThrTrpTyrAlaIleGlyThrValAlaSerMetGln 252
Db 879 ATGAGCTTTTACATGCTTATTAACGTATGATATATCTGGGGATGTTGTCGCGATCAA 938
QY 253 IleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAlaGlyVal 272
Db 939 ATTTCGCTTCTGTTCTTC---ACATGTTTCATCTGGTGAACATATGGGTGCATGGGTGTC 995
QY 273 PheGlyLeuIleGlnIleValAlaPheGlyAspPheValLysGlyGlnIleSerThrAla 292
Db 996 TTCTTATTATGAGGCTGTTTCTTCTTCTGACTGGGTAAAGTACCAGCTCAATGACACC 1055

```


QY 293 Lys-...PheLysValIleMetMetValSerLeuPheLeuIleLeuValLeuGlyValVal 311
 Db 1056 AAGTTGTCACAAACGTTTTGAGATAAATCGGCTATTCGTGGTGGT 1115
 QY 312 GlyLeuSerAlaLeuThrThrMetGlyLeuIleAlaProThrGlyArgPheThrSer 331
 Db 1116 GCTGGGGTGTGGACACGCTCTGGCTATATCTCCATGAGCTGGTGCATTTACTCG 1175
 QY 332 LeuTrpAspThrAsnTrpAlaLysIleHisIleProIleIleAlaSerValSerGluHis 351
 Db 1176 TTGCTTGACCGCACTTACGCAAGGATCATATTCCTATTATGATCTGTCTTCTGAGCAT 1235
 QY 352 GlnProValSerTrpProAlaPhePhePheThrHisPheLeuIleTrpLeuPhePro 371
 Db 1236 CAGCTTACTGCTGGTGTGCTATCATGTTGCTACCATATTCGCTTCTTCTTCCCT 1295
 QY 372 AlaGlyValPheLeuLeuPheLeuAspLeuLysAspGluHisValPheValIleAlaTrp 391
 Db 1296 GCGGGTCTTACTTCTGTTTCAAGGCTTTCAGGATGCTAGCATATTTATGTCATGAT 1355
 QY 392 SerValLeuCysSerTrpPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPro 411
 Db 1356 GGTCTCACTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
 QY 412 ValIleCysValSerAlaAlaValAlaLeuSer-----LysIlePheAspIleTrp 428
 Db 1416 GCAGTTGTCTTATCAGTGCCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1475
 QY 429 LeuAspPheLys----- 432
 Db 1476 TTAAGCAAAACAAAGGTTTCTCAGACTGTTCCAGAAAGAGCTGTGATGTCAAAA 1535
 QY 433 -----ThrSerAspArgLysTrpAlaIleLysProAlaAlaLeuLeuAla 1547
 Db 1536 GCTTCTTCAAGGTTACACTTGCATCAGTCTGAGCTTCCAGAGAGATGGTGCCTGCT 1595
 QY 448 LysLeuIleValSerGlySerPheIlePheTrpLeuTrp-----LeuPheValPheHisSer 466
 Db 1596 CTTCTTGT-----GGTGATTTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1643
 QY 467 ThrTrpValThrArgThrAlaTrpSerSerProSerValValLeuProSerGluThrPro 486
 Db 1644 ACATGGGTGACAGCAGGATATTCATCTCTCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1703
 QY 487 AspGlyLysLeuAlaLeuIleAspPheArgGlyAlaTrpTrpLeuArgMetAsn 506
 Db 1704 CACGGGACAGAAATCATCTTTGATGATACCGGAGGCTATCTGCTGCTGCTGCTGCTGCTGCTGCT 1763
 QY 507 SerAspGluAspSerLysValAlaAlaTrpTrpAspTrpGlyTrpGlyGlyGlyMet 526
 Db 1764 ACTGCCACTGATGCTAAGATTATGCTGGTGGGACTATGGGTACCAAAATCCTGCTGCTGCTGCT 1823
 QY 527 AlaAspArgThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleVal 546
 Db 1824 GGAATAGAACTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1883
 QY 547 GlyLysAlaMetAlaSerProGluGluLysSerTrpGluIleLysGluHisAspVal 566
 Db 1884 GGACGCCCATGCTTCTTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1943
 QY 567 AspTrpValLeuValIlePheGlyLeuIleGlyPheGlyGlyAspIleAsnLys 586
 Db 1944 AACTATGATTTGGTGTGCTTGGTGTGCTTGGTGTGCTTGGTGTGCTTGGTGTGCTTGGTGTGCTTGGT 2003
 QY 587 PheLeuTrpMetIleArgIleSerGluGlyIleTrpProGluGluIleLysGluArgTrp 606
 Db 2004 TTTCTTGGATGGTGGAGATTGGAGTGGAGTATTTCCCA---GTGATCAAGAACCT--- 2057
 QY 607 PheTrpThrAlaGluGlyGluTrpArgValAspAlaArgAlaSerGluThrMetArgAsn 626
 Db 2058 GATTACCTTGTGAATGGTGAATTCCTGTAGACAAAGGTGATCACCAGAGATGCTGAAC 2117

QY 627 SerLeuLeuTrpLysMetSerTrpLysAspPheProGlnLeu-----PheAsnGly 643
 Db 2118 TGCTCTCATGTACAAAGTTATGCTATTACAGGTTCGGTGTGCTGACCAAGATACGCAAG 2177
 QY 644 GlyGlnAlaThrAspArgValArgGlnGlnMetIleThrProLeuAspValProProLeu 663
 Db 2178 CCACCCGGGTATGATAGCAAGAGGAGTGGAGTATGGCAACAAAGATATC---AAACTT 2234
 QY 664 AspTrpPheAspGluValPheThrSerGluAsnTrpMetValArgIleTrpGlnLeuLys 683
 Db 2235 GAACACTTAGAAGAGCTTACACAACTCCAACTGAGTATGCTGCTATTTACAGAGTAAA 2294
 RESULT 12
 AY418285 2079 bp DNA linear GSS 17-DEC-2003
 LOCUS Pan troglodytes ITMI gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY418285
 VERSION AY418285.1 GI:39774245
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 (bases 1 to 2079)
 AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 2079)
 AUTHORS Clark,A.G., Glancowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (18-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 Location/Qualifiers
 1..2079
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 gene
 <1..>2079
 /gene="ITM1"
 /locus_tag="HQM6506"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,48e-152 Length: 2079
 Score: 1465.00 Matches: 304
 Percent Similarity: 57.43% Conservative: 71
 Best Local Similarity: 46.55% Mismatches: 246
 Query Match: 38.95% Indels: 32
 DB: Gaps: 7
 US-10-028-384-6 (1-718) x AY418285 (1-2079)
 QY 50 PheAsnTrpArgAlaThrLysTrpLeuValAsnAsnSerPheTrpLysPheLeuAsnTrp 69
 Db 154 TTTAATTATCGACTACCAAGNNCTGCTGNNNGGGTTTATAAATCCATACTGG 213
 QY 70 PheAspAspArgThrTrpTrpProLeuGlyArgValThrGlyThrLeuTrpProGly 89
 Db 214 TTGTGACCCAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 273
 QY 90 LeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIle 109
 Db 90 LeuMetThrThrSerAlaPheIleTrpHisAlaLeuArgAsnTrpLeuGlyLeuProIle 109

| | | |
|------|--|--------------------------|
| 274 | TTAAGATCACTCTGCTGCAATCTACCAATGACTC--- | CATTTTTCCACATCACCATC 330 |
| 110 | AspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAla 129 | |
| 331 | GACATTCGGAATGCTNTGTGTTCTCTGGCCCTCTCTTCTCTCTCTCCACACCATCGTC 390 | |
| 130 | ThrTyrGluPheThrIysGluIleIysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPhe 149 | |
| 391 | ACGTACCACTTACCAAGAGCTCAAGGATCAGGGCTGGCGCTCTCTCTCTCTGCCATG 450 | |
| 150 | IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAla 169 | |
| 451 | ATTGCTGTAGTTCCTCGATATATCTCCGATCTGTGGCTGCTCTATGATAATGAAGGG 510 | |
| 170 | IleAlaIleThrLeuLeuMetValThrPheMetPheTyrIleIysAlaGlnIleThrGly 189 | |
| 511 | ATTGCCATCTTTGCATGCTACTCACTTACTATGTGTGATCGATCAGGCGATTAAGACTGGT 570 | |
| 190 | SerIleMetHisAlaThrCysAlaAlaLeuPheTyrPheTyrMetValSerAlaTrpGly 209 | |
| 571 | TCCATCTGTTGGGAGCTAAGTGTGCCCTTGCTTATTCTATCATGGTCTGTCATGGGA 630 | |
| 210 | GlyTyrValPheIleThrAsnLeuIleProLeuHisValPheLeuLeuIleLeuMetGly 229 | |
| 631 | CGTTATGTGTTCTGTGATCACTTAATCTCTCTCCAGCTCTCGTGTGATGCTCAGGC 690 | |
| 230 | ArgTyrSerSerIysLeuTyrSerAlaTyrThrTyrAlaIleGlyThrValAla 249 | |
| 691 | CGTTTCTCTCACCGGATCATGTGGCTACTGTACTGTATTCTGCTGGGCACATATACT 750 | |
| 250 | SerMetGlnIleProPheValGlyPheLeuProIleArgSerAsnAspHisMetAlaAla 269 | |
| 751 | TCTATGAGATCTCCTTTGTGGGTTTCCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 810 | |
| 270 | LeuGlyValPheGlyLeuIleGlnIleValAlaPheGlyAspPheValIysGlyGlnIle 289 | |
| 811 | NN 870 | |
| 290 | SerThrAlaIysPheIysValIleMetMetValSerLeuPheLeuLeuValLeuGly 309 | |
| 871 | AATCCAAACAATTTNNNGTCTTTCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 930 | |
| 310 | ValValGlyLeuSerAlaLeuThrTyrMetGlyLeuIleAlaProTyrThrGlyArgPhe 329 | |
| 931 | NN 990 | |
| 330 | TyrSerLeuTrpAspThrAsnTyrAlaIysIleHisIleProIleIleAlaSerValSer 349 | |
| 991 | TACTCACTGTGGATCCTTCTATGCTAAGAACACATCCCATCATGTTGTTCTGTCTCT 1050 | |
| 350 | GluHisGlnProValSerTrpProAlaPhePheAspThrHisPheLeuIleTrpLeu 369 | |
| 1051 | GAGCATCAGGCCACACCTGGTCTCATATATTGTACCTTNNAGCTCTCTGCTTCATG 1110 | |
| 370 | PheProAlaGlyValPheLeuLeuPheLeuLysAspGluHisValPheValIle 389 | |
| 1111 | TTTCCAGTGGCCTCTATTACTGCTTTTGCAACCTGCTGTGATGCCGGATTTTATCATC 1170 | |
| 390 | AlaTyrSerValIleuCysSerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 409 | |
| 1171 | ATGTATGGTGTGACAGCATGTACTTTTCAGCTGTAATGNNNNNNNNNNNNNNNNNNNN 1230 | |
| 410 | ThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeu 429 | |
| 1231 | NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGAGTCTCCAGGTTNTGTCNNNATACATG 1290 | |
| 430 | AspPhe-----LysThrSerAspArgLysTyr 438 | |
| 1291 | AAGAANTGACATAANNNNNNNNNNNAGACCAAGACAGANNNNNAGGATTCGCCCNCC 1350 | |
| 439 | AlaIleLysProAlaLeuLeuAlaLysLeuIleValSerGlySerPheIlePheTyr 458 | |
| 1351 | CCTATTGAAGATGAAGTGGCAAGTGGGATGATATGGTGC-----ATGGCTTCTTCTT 1401 | |

459 LeuTyLeuPheValPheHisSerThrTrpValThrArgThrAlaTyrSerProSer 478
 1402 CTATCACCTACACCTTTTCATTCAACCTGGGTGACCATGGAGCCCTACTCTTCTCGTGCC 1461
 479 ValValLeuProSerGlnThrProAspGlyLysLeuAlaLeuAlaAspAspPheArgGlu 498
 1462 ATTGTACTATCTGCCCGTGGTGGGATGGCAGTAGGATCATATTTCGTGACTTCGCAGAA 1521
 499 AlaTyrTyTrpLeuArgMetAsnSerAspGluAspSerLysValAlaAlaTrpTrpAsp 518
 1522 GCATATTATTGGCTTCGTCCAAATACTCCAGAGATCGAAGGTCATNNCCTGTGTGGGAT 1581
 519 TyrGlyTyGlnIleGlyGlyMetAlaAspArgThrThrLeuValAspAsnAsnThrTrp 538
 1582 TATGGCTATCAGATTACAGCTATGGCAACCGGAACAATTTAGTGGACAATAACACATGG 1641
 539 AsnAsnThrHisIleAlaIleValGlyLysAlaMetAlaSerProGluGluLysSerTyr 558
 1642 AATAATACCATATTCTCINNNNNNNNGCATGGCGTCCACAGAGAGAAAAGCCTAT 1701
 559 GluIleLeuLysGluHisAspValAspTyrValLeuValIlePheGlyLysLeuIleGly 578
 1702 GAGATCATGGAGGAGCTCGATGTCAGCTATGCTGTGTCATTTTGGAGCCTCACTGGG 1761
 579 PheGlyGlyAspAspIleAsnLysPheLeuTrpMetIleArgIle---SerGluGlyIle 597
 1762 TATTCTCTGATGATATCAACAAGTTCTNNNGATGGTCCGGATGGAGGGAGNACAGAN 1821
 598 TrpProGluGluIleLysGluArgTyrPheTyThrAlaGluGlyGluTyArgValAsp 617
 1822 NNANGCAACATATTCAGGAGAGATGACTATTATTACTTCAACTGGGAGTTCGTGTGNNN 1881
 618 AlaArgAlaSerGluThrMetArgAsnSerLeuLeuTyrLysMetSerTyrLysAspPhe 637
 1882 CNTGAAGGTTCTCCAGTGTCTGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1941
 638 -----ProGlnLeuPheAsnGlyGlyGlnAlaThrAsp 648
 1942 NNNNNNNNNNNNNNNNNNNNAGGCTCTCCAGGCTTT-----GAC 1983
 649 ArgValArgGlnGlnMetIleThrProLysAspValProProLeuAspTyrPheAspGlu 668
 1984 CGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTT---GAGCTTGATGCTCTGGAGGAA 2040
 669 ValPheThrSerGluAsnTrpMetValArgIleTyGln 681
 2041 GCATATACACAGAAATTTGGTGTGCTCAGGATATACAA 2079

RESULT 13

| | | | | | |
|------------|---|-------------|-----|--------|----------------|
| CNS06CLU/c | CNS06CLU | 1088 bp | DNA | linear | GS 14-JUN-2001 |
| LOCUS | T7 end of clone AF0A006C05 of library AROAA from strain CBS 732 of | | | | |
| DEFINITION | Zygosaccharomyces rouxii, genomic survey sequence. | | | | |
| ACCESSION | AL392680 | | | | |
| VERSION | AL392680.1 | GI:12142260 | | | |
| KEYWORDS | GS. | | | | |
| SOURCE | Zygosaccharomyces rouxii | | | | |
| ORGANISM | Zygosaccharomyces rouxii | | | | |
| | Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | | |
| | Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes. | | | | |
| REFERENCE | 1 (bases 1 to 1088) | | | | |
| AUTHORS | Soucet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Boilotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durren,P., Lepingle,A., Morente,B., Maurituy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wanker,P. and Weissenbach,J. | | | | |
| TITLE | Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies | | | | |
| JOURNAL | FEBS Lett. 487 (1), 3-12 (2000) | | | | |
| MEDLINE | 20584711 | | | | |
| PUMLID | 11152876 | | | | |


```

646 -----TACAGGGTCTCTGGGAGGAGGTTATGTGTTTCATCATCAACCTCATCCCT 696
220 LeuHisValPheLeuLeuLeuLeuMetGlyArgTyrSerSerLysLeuTyrSerAlaTyr 239
697 CTCCTATGTTTGTGCTGCTCTCTGATGAGAGGACACAGAGAGAGTCTAC----- 747
240 ThrThrTyrAlaAlaLeuGlyThrValAlaSerMetGlnIleProPheValGlyPheLeu 259
747 ----- 747
260 ProIleArgSerAsnAspHisMetAlaAlaLeuGlyValPheGlyLeuLeuGlnIleVal 279
748 -----ATAGGTGCTTTTGGCTGCTGCACGCTTAT 777
280 AlaPheGlyAspPheValGlyGlnIleSerThrAlaLysPheLysValIleMetMet 299
778 GCGTTTCTCCAGTATCTGAGAGACCGGTTGACAAAGCAGGAGTCCAGACCTTTTCTTT 837
300 ValSerLeuPheLeuLeuValLeuGlyValValGlyLeuSerAla-----LeuThr 317
838 -----TTGGGTGCTCTACTAGCTGAGCGGCTGTGTTCTTAGTGTCTATCTATCTGACA 891
318 TyrMetGlyLeuLeuAlaProThrGlyArgPheTyrSerLeuTyrAspThrAsnTyr 337
892 TAC-----ACAGG-TAT 902
338 AlaLysIleHisIleProIleAlaSerValSerGluHisGlnProValSerTrpPro 357
903 GCAAAATATACATCCAAATATTGTCATCAGTGTCTGAACATCAGCTCAGCATGGGTG 962
358 AlaPhePheAspThrHisPheLeuLeuTyrLeuPheProAlaGlyValPheLeuLeu 377
963 TCTTTCTCTTGAICTACATATTCTGTGTGTACTCTCCAGCAGGCTATGGTCTTGC 1022
378 PheLeuAspLeuLysAspGluHisValPheValIleAlaTyrSerValLeuCysSerTyr 397
1023 ATCAAAATATACAGTGAAGAGTATT----- 1052
398 PheAlaGly-ValMetValArgLeuMetLeuThr-----le 409
1053 -----GGCAGGGAAGTGAAGAGCATGTGACGGAGCAGGAGAGCCGAGGAGGCTT 1106
409 uThrProValIleCysValSerAlaAlaValAlaLeuSerLysIlePheAspIleTyrle 429
1107 GGCCCAACATC----- 1119
429 uAspPheLysThrSerAspArgLysTyrAlaIleLysProAlaAlaLeuLeuAlaLysLe 449
1120 -----AAGAG 1124
449 uIleValSerGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSerThrTyrVa 469
1125 CATCGTACCATCTCATGCTCATGCTCTCATGATGTTCTGCTGTCATCGACCTGGGT 1184
469 lThrArgThrAlaTyr-SerSerProSerValValLeuProSerGlnThrProAspGlyLys 489
1185 CACCAAGAACCCCTACTCCAGCCCCAGTGTGTCTCTGCTCTCAATCATGATGTAAC 1244
489 sLeuAlaLeuIleAspAspPheArgGluAlaTyrTyrTyrLeuAtgMetAsnSerAspGl 509
1245 CAGAAATATCTAGATGATTTAGAGAGCATCTTTGTTGAGACAAACACCGATGA 1304
509 uAspSerLysValAlaAlaTyrTrpAspTyrGlyTyrGlnIleGlyLysMetAlaAspAr 529
1305 ACAGCCAGGCTCATGCTCTGTTGGGACTATGCTATCATGATGTTGGTGGTGGCAACAG 1364
529 gThrThrLeuValAspAsnAsnThrTrpAsnAsnThrHisIleAlaIleValGlyLysAl 549
1365 GACTACTCTGATAGCAATTAACACCTGGAAACACAGCCATTCGCTAGTGGAAAGC 1424
549 aMetAlaSerProGluGluLysSerTyrGluIleLeuLysValAspValAspTyrVa 569
1425 TATGTCTCTTCTAATGAACACCCGCTTATAAAATCATGAGGTCTCTTGATGTTATGT 1484

```

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
232-2334
CDS
/feature="liver regeneration related protein LRRG109"
/codon_start=1
/product="Ac1573"
/protein_id="AAP78765.1"
/db_xref="GI:32264637"
/translation="MTITAGLHILNLTINITVLRDVCVFLAPTFSLGTLSTIFLLTR
ELWQAGALLACPIAIVPGVYSRSVAGSPNEGIAIFALQFTVYLMHSGVGLCPRA
HRSPFGSSGSLTISVAALPSGALVHVARPELYRVSAGVYFIINLIPIHVFV
LLLMRYKRVYIGVAFLLQAYAFQYLRDLTKQERQTFLLFGVSLAAGAFSLVY
LTYTQMOKYTLQHLQCLNLSRHGCLSSLIYFLCVPQOAGSASKISTMKEYLAG
KVRHTVQKPEEGLGNIKSIIVTLMMLMFAVCTWTTSNAYSSPVSLASYN
HYRTNLDLDFREAFVLRQNTDEHARVMSWDYQYIAGMANRTLLVDNNTWNNSHI
ALVGKMSNSTEAKVIRSLDVLVIFGVVIGYSGDDINKFLMVRVLAEGEHPKD
IRGCVFTQCEFFRVDKAGSPTLNCIMYKMSYVRFGEMLDRTPPGFDRTNABIG
NKDIKFLAEAFTEHHLVYIKVAPDRNETLGHKPRVTNIPPKOKYLSKKDLAAS
TEERPGHTQGLTVSEEQKRWRLSVSGILGSEENVLPDCCQHSHSDHNLHQEQQ
MNTDLGLGSQVSAKGSFPQDRAPFSDRDNFLCPFLHMDMSNQNCNVRKNT
ESREQKAQDPGPPRV"
ORIGIN
Alignment Scores:
Pred. No.: 2,48e-145 Length: 2334
Score: 1404.50 Matches: 301
Percent Similarity: 57.42% Conservative: 90
Best Local Similarity: 44.20% Mismatches: 144
Query Match: 37.34% Indels: 147
DB: 3 Gaps: 12
US-10-028-384-6 (1-718) x AY310157 (1-2334)
QY 50 PheAsnTyrArgAlaThrLysTyrLeuValAsnAsnSerPheTyrLysPheLeuAsnTrp 69
DB 109 TTTAACTATAGATCACTCATCTTGTGATCTGTGCTTCTATGAGTCTTAAATGG 168
QY 70 PheAspAspArgThrTrpTyrProLeuGlyArgValThrGlyGlyLeuTyrProGly 89
DB 169 TTGTGAAAGAGCATGGTACCACCTAGGAAGATCTGGTGGTGGCCCGTTTACCAGGA 228
QY 90 LeuMetThrThrSerAlaPheIleThrHisAlaLeuArgAsnTrpLeuGlyLeuProIle 109
DB 229 TTGATGATACAGCTGCGCTTATT--CATTGATTTTAAATACATTGAACATACAGTT 285
QY 110 AspIleArgAsnValCysValLeuPheAlaProLeuPheSerGlyValThrAlaTrpAla 129
DB 286 CACATCAGAGATGATGTGTGTTCTGTGACCAACTTTTAGCGGCTTACATCTATATCT 345
QY 130 ThrTyrGluPheThrLysGluIleLysAspAlaSerAlaGlyLeuLeuAlaAlaGlyPhe 149
DB 346 ACCTTCTCTACTAGAGACTTTTGGAAACAGGAGCAGGCTCTTAGTCTGCTGCTC 405
QY 150 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluAla 169
DB 406 ATTGCTATTGTATCCCGGATACATATCTCGTCCGTAGCGGATCCCTTTGATATAACAGGC 465
QY 170 IleAlaIleThrLeuLeuMetValThrPheMetPheTrp----- 182
DB 466 ATTGCGATTTCGCATCTTCAGTTCACTTACTTATGGGCACTTCTGGGCTTCTGCTC 525
QY 182 ----- 182
DB 526 TGCTTTCTGCCCATCGCTCCAGGCCCCCTTTGGTTCTCAGGGCACTTCACTCACCATC 585
QY 183 -----IleLysAlaGlnLysThrGlySerIleMetHisAlaThrCysAlaAlaLeu 199
DB 586 TCCTGTGTGGCCGCTGCTCTTCTAGTGGGCGCTGATACATGTTGCTCGGCTGAGCTC 645
QY 200 PheTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleThrAsnLeuLeuPro 219

```

QY 569 lieuValliePheGlyLeuileGlyPheGlyAspAspIleAsnIlySphLeuYr 589
 Db 1495 GTTGTCTATTTTGGAGGATTTATGGTCTTCTGGTATGATATCAACAGTTCTCTG 1544
 QY 589 pMetIleArgIleSerGluGlyIleTrrProGluGlyIleLysGluArgTyrPheYrTh 609
 Db 1545 GATGTCAGATAGCTGAGGGAGACCCAAAGACATTCGGAGGCGNACTTTCAC 1604
 QY 609 rIaGluGlyIleTrrArgValAspAlaArgAlaSerGluThrMetArgAsnSerLeuLe 629
 Db 1605 CCAGCAGGAGGATTTCCGAGTGGACAGAGCCGGTCTCTACTCTGTAAATTCCTTAT 1664
 QY 629 uYrIlySMetSerTyrLysAspPheProGlnLeu-----PheAsnGlyGlyGlnAl 646
 Db 1665 GTATAAATGTCATCTACTACAGATTGGAGAAATGACGACTAGTTTGCACACCCCGG 1724
 QY 646 aThrAspArgValArgGlnGlnMetIleThrProLeuAspValProLeuAspTyrPh 666
 Db 1725 TTTTACCAGAACACGTAATGCTGAGATCGGAATAAGGACATT---AAATTCAGCATTT 1781
 QY 666 eAspGluValPheThrSerGluAsnTrrMetValArgIleTyrGlnLeuIlySlySAspAs 686
 Db 1782 CGAGGAGGACCTTTATCATGACGACTGGCTGTGAGGATATATAAGTGAAGACCCCGA 1841
 QY 686 p 686
 Db 1842 C 1842

RESULT 15
 CNS06YOT
 LOCUS
 DEFINITION T3 end of clone AY0AA005C11 of library AY0AA from strain CBS 6340
 ACCESSION AL420435
 VERSION AL420435.1 GI:12203620
 KEYWORDS GSS.
 SOURCE Kluyveromyces thermotolerans
 ORGANISM Kluyveromyces thermotolerans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 1 (bases 1 to 981)
 Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
 Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
 de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,
 Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S.,
 Saurin, W., Tekais, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
 Wincker, P. and Weissenbach, J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 20594711
 PUBMED 11152876
 REFERENCE
 AUTHORS Malpertuy, A., Llorente, B., Blandin, G., Artiguenave, F., Wincker, P.
 and Dujon, B.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 10.
 Kluyveromyces thermotolerans
 FEBS Lett. 487 (1), 61-65 (2000)
 20594720
 PUBMED 11152885
 REFERENCE
 AUTHORS 3 (bases 1 to 981)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 TITLE This GSS is part of a random genomic sequencing program of thirteen
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
 angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source
 1. 981
 /organism="Kluyveromyces thermotolerans"
 /mol_type="genomic DNA"
 /strain="CBS 6340"
 /db_xref="taxon:4916"
 /clone="AY0AA005C11"
 /clone_id="AY0AA"
 /note="Tend : T3"
 <7..>939
 misc_feature
 /note="similar to Saccharomyces cerevisiae ORF YGL022w [S113 ; oligosaccharyl transferase subunit]"
 /evidence=not_experimental

ORIGIN

Alignment Scores:
 Pred. No.: 1303-134 Length: 981
 Score: 1303.50 Matches: 250
 Percent Similarity: 84.71% Conservative: 27
 Best Local Similarity: 76.43% Mismatches: 41
 Query Match: 76.46% Indels: 10
 DB: 9 Gaps: 2

US-10-028-384-6 (1-718) x CNS06YOT (1-981)

QY 336 AntTyrAlaIlyIleHisIleProIleAlaSerValSerGluHisGlnProValSer 355
 Db 1 AACACAGGAGATCCACATCCCATATATCGCTCGGTATCAGAGACACGCTACCGCG 60
 QY 356 TrrProAlaPhePhePheAspThrHisPheLeuIleTrrPhePheProAlaGlyValPhe 375
 Db 61 TGGCCAGGCTTTTCTTTGATGATCTCAATTTCTGATTTGGTTATTTCCAGCGGGTATTT 120
 QY 376 LeuLeuPheLeuAspLeuHisValPheValIleAlaTyrSerValLeuCys 395
 Db 121 TTTCTGTTCTTGGATCTCAAGATGAACAGCTTTCTGATTTCTTCTCTCTCTGTC 180
 QY 396 SerTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValIleCysVal 415
 Db 181 TCTTATTTTGGGGTGTGTCATGGTCAGACTCATGTAATCTGACACCAATCATCTGTGTC 240
 QY 416 SerAlaAlaValAlaLeuSerLysIlePheAspIleTyrLeuAspPheLys---ThrSer 434
 Db 241 TGTGCGGAGTAGGCAATTTCCAGCTATTCAGCTCTACGCGCTTTCATATTTTTC 300
 QY 435 AspArgLysTyrAlaIleLysPro-----AlaAlaLeuLeu 446
 Db 301 AGTTCTCCAGCGCAGAAAGCCAAACCTGTCCAAGAGACGCGCATTTGCTTTGGTT 360
 QY 447 AlalysLeuIleValSerGlySerPheIlePheTyrLeuTyrLeuPheValPheHisSer 466
 Db 361 ACTAAGCTAACGGTTTGTGGTTCTTCTTGGCCCTATTGTTTGTGCGGCACCACTGT 420
 QY 467 ThrTrrPValThrArgThrAlaTyrSerSerProSerValValLeuProSerGlnTrrPro 486
 Db 421 ACCTGGGTGACCTCTAACGCTTATTCTGCGCATTCGTTGGTTTTCCTTCACGTAATCC 480
 QY 487 AspGlyLysLeuAlaLeuIleAspPheArgGluAlaTyrTrrPleuArgMetAsn 506
 Db 481 GACGATCTCTGCTCTGATAGACGACTTCGGTAGGCTTACTACTGCGGTATGAAC 540
 QY 507 SerAspGluAspSerLysValAlaAlaTrrTrrAspTyrGlyTyrGlnIleGlyGlyMet 526
 Db 541 ACTGACAGAGATGCCAGGTTGCTCTTGTGGGACTACGCTACCAATTTGTTGGCATG 600
 QY 527 AlaAspArgThrThrLeuValAspAsnThrTrrPheAsnThrHisIleAlaIleVal 546
 Db 601 GCTGACGAACACCTGGTGTAGAACAACTTGGAAACCAACCCACCATCGCTATCGTG 660

```
QY 547 GlyLysAlaMetAlaSerProGluGluLysSerTyrGluIleLeuLysGluHisAspVal 566
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 GGTAAGCTATGTCGTCCTCCCAAGAGAGAGCGGTATGACATCTTGAAGGCTCAGCATGTG 720
QY 557 AspTyrValLeuValIlePheGlyGlyLeuIleGlyPheGlyGlyAspIleAsnLys 586
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
721 GACTACGTTCTGGTTATTTTCGGCGGGCTCATAGGATTCTCAGGTGATGATATCAACAAG 780
QY 587 PheLeuTyrMetIleArgIleSerGluGlyIleTyrProGluGluIleLysGluArgTyr 606
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
781 TTCCTATGGATGGTAAGGATTTCTGAGGGCATATGCCCGCATGAATCAAGAGAGAGAC 840
QY 607 PheTyrThrAlaGluGlyGlyTyrArgValAspAlaArgAlaSerGluThrMetArgAsn 626
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
841 TACTTCACCCCGGCGGTGAGTACCGGTATGCGCGCAACTDTCCTCCAAACCATGAAGAC 900
QY 627 SerLeuLeuTyrLysMetSerTyrLysAspPheProGluLeuPheAsnGlyGlyGlnAla 646
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 TCGCTGTGTGTACAGCTCTCTACCCGCTTCCGACA-TTGTTCATGGSGTTGAGGCG 959
QY 647 ThrAspArgValArgGlnGln 653
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
960 ATCAGCCGCTCTGCGTGTCTCAG 980
```

Search completed: December 15, 2004, 11:51:40
Job time : 5663.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 02:41:23 ; Search time 16231.6 Seconds
(without alignments)
7041.770 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctagcgaagaatgtgctg.....ccaaaaaAAAAAAAAAAAA 2417

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------------------|
| 1 | 2417 | 100.0 | 2417 | 3 | AF132552 Drosophil |
| 2 | 2417 | 100.0 | 2417 | 6 | AX799088 Sequence |
| 3 | 2409 | 99.7 | 2899 | 6 | CQ589353 Sequence |
| C 4 | 1463 | 60.5 | 4922 | 6 | CQ589352 Sequence |
| C 5 | 1463 | 60.5 | 162321 | 3 | AC007853 Drosophil |
| C 6 | 1463 | 60.5 | 181132 | 3 | AC008206 Drosophil |
| C 7 | 1463 | 60.5 | 227219 | 3 | AE003750 Drosophil |
| C 8 | 1342 | 55.5 | 75650 | 2 | AC018145 Drosophil |
| C 9 | 696 | 28.8 | 2953 | 6 | CQ596819 Sequence |
| C 10 | 249 | 10.3 | 2785 | 6 | CQ596795 Sequence |
| 11 | 29 | 1.2 | 297 | 8 | AF548377 Talaromyc |
| 12 | 29 | 1.2 | 218635 | 2 | AC118338 Rattus no |
| 13 | 29 | 1.2 | 263377 | 2 | AC105647 Rattus no |
| 14 | 29 | 1.2 | 285045 | 2 | AC116217 Rattus no |
| 15 | 28 | 1.2 | 1968 | 9 | AK074363 Homo sapi |
| 16 | 28 | 1.2 | 110000 | 2 | PF04113-26 Continuation (27 o |
| C 17 | 28 | 1.2 | 196966 | 10 | AC087898 Mus muscu |
| C 18 | 28 | 1.2 | 269583 | 2 | AC11344 Rattus no |
| C 19 | 27 | 1.1 | 391 | 6 | CQ433488 Sequence |

| | | | | | |
|------|----|-----|--------|----|--------------------------------|
| C 20 | 27 | 1.1 | 402 | 6 | CQ415729 Sequence |
| C 21 | 27 | 1.1 | 471 | 6 | CQ424280 Sequence |
| C 22 | 27 | 1.1 | 1420 | 3 | AY513651 Ostrinia |
| C 23 | 27 | 1.1 | 1420 | 3 | AY513652 Ostrinia |
| C 24 | 27 | 1.1 | 2158 | 5 | BC061422 Xenopus t |
| C 25 | 27 | 1.1 | 2270 | 9 | BC041156 Homo sapi |
| C 26 | 27 | 1.1 | 2474 | 9 | BSM806789 Homo sapi |
| C 27 | 27 | 1.1 | 156070 | 2 | AC110726 Mus muscu |
| C 28 | 27 | 1.1 | 173871 | 2 | AC102469 Mus muscu |
| C 29 | 27 | 1.1 | 180998 | 2 | AC113766 Rattus no |
| C 30 | 27 | 1.1 | 185342 | 2 | CR387989 Dario rer |
| C 31 | 26 | 1.1 | 786 | 10 | BC037513 Mus muscu |
| C 32 | 26 | 1.1 | 1332 | 8 | AF370133 Arabidops |
| C 33 | 26 | 1.1 | 1888 | 3 | AX113753 Ciona int |
| C 34 | 26 | 1.1 | 2855 | 6 | CQ573675 Sequence |
| C 35 | 26 | 1.1 | 3248 | 6 | AX753657 Sequence |
| C 36 | 26 | 1.1 | 3418 | 9 | BC037539 Homo sapi |
| C 37 | 26 | 1.1 | 4822 | 10 | BC064459 Mus muscu |
| C 38 | 26 | 1.1 | 6153 | 6 | CQ573674 Sequence |
| C 39 | 26 | 1.1 | 9765 | 3 | AY160106 Dictyoste |
| C 40 | 26 | 1.1 | 18405 | 2 | AC017914 Drosophil |
| C 41 | 26 | 1.1 | 104992 | 2 | AC005504 Plasmodiu |
| C 42 | 26 | 1.1 | 110000 | 8 | CE382128 21 Continuation (22 o |
| C 43 | 26 | 1.1 | 114136 | 10 | AL586447 Mouse DNA |
| C 44 | 26 | 1.1 | 161577 | 2 | AC118924 Mus muscu |
| C 45 | 26 | 1.1 | 166908 | 3 | AC104631 Drosophil |

ALIGNMENTS

| | | | | | | |
|------------|---|-------------------------|---------|-------------|--------|-----------------|
| RESULT 1 | AF132552 | Drosophila melanogaster | 2417 bp | mrna | linear | INV 12-FEB-2003 |
| LOCUS | AF132552 | Drosophila melanogaster | GM01838 | full insert | cdna. | |
| DEFINITION | AF132552 | Drosophila melanogaster | GM01838 | full insert | cdna. | |
| ACCESSION | AF132552 | Drosophila melanogaster | GM01838 | full insert | cdna. | |
| VERSION | AF132552.1 | GI:4689327 | | | | |
| KEYWORDS | FLI CDNA. | | | | | |
| SOURCE | Drosophila melanogaster | (fruit fly) | | | | |
| ORGANISM | Drosophila melanogaster | | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | Tsang, G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3200, Berkeley, CA 94720, USA | | | | | |
| REFERENCE | 2 | (bases 1 to 2417) | | | | |
| AUTHORS | Stapleton, M., Brokstein, P., Hong, L., Agbavani, A., Carlson, J., Champagne, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nuncio, J., Pacleb, J., Pargass, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Celniker, S. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (12-FEB-2003) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA | | | | | |
| COMMENT | Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory Berkeley, CA 94720 This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal | | | | | |

priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

Source

1..2417
Location/Qualifiers

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

<1..2417

/gene="OstStt3"

/db_xref="FLYBASE:FBgn0011336"

<1..2395

/gene="OstStt3"

/note="Longest ORF"

/codon_start=2

/product="GM01838p"

/protein_id="AAD27851.2"

/db_xref="GI:28317200"

/db_xref="FLYBASE:FBgn0011336"

/translation="LSRECVAFQIGNVPRVTVGNDNRTPKMLNSKVAGYSSLIITP
ALLIALLACFSSRLFAVIREPSLIHEEPFNXYRATAYMVGWYNFLNPNFDEBAY
ELRIVGTVYPLGMITSGGIHMLLVNIPVHIRDICVFLAPISGTSISTYLTK
LWAGAGLFAASFIAPVGYIISRVAGSYDNEGIAIFALQTYIFLWVRSVKTSVFW
SAAALSYFVMSWAGYVFIINLIPLHVFLLIMRGYSRPLLTYSYFYLGLFSM
QIPVGPQIRTSHEMALGVFLLMAVATLRLQSVLRNEFRKLFYVGLLVGVV
FVAVVITMLGVAPWSGRFYSLMDTGYAKIHPIIASVSEHQPWTWFFDFDLHLV
CAFPVGVYCIKQINDRERFVLYAISAVYFAGVWVRLMLTTPVVCMLAGVAFGLL
DVLQEDSSKRMGTASAEVDEBARSIEKTLXDKAKLKHRTKHDAQDDTVGSN
LKSIVILVALLMFWVHCTWTSNAYSPSIVLAFHNSODGSNILLDEREAYML
SCNTADARVWSWDYGYQIAGMANRTLLVDNNWNHSHLVKAMSTSEKSHYM
ISLDVILVIFGVGIVSGDINKLFWIRAEGBPKDINKESDYFDIRGEFRVDAE
GAPALLNLYKUSYFBEKLDYRPGSYDRNNAVIGNKDFDLTLEEAYTTEHW
LVRIYRVKPHFENRPSLKTERTIPPANFISFKNSRRKRYIRNRPVVVKGRKTLK"

ORIGIN

| | | | | |
|-----------------------|-----------------|---------------|-----------|--------------|
| Query Match | 100.0%; | Score 2417; | DB 3; | Length 2417; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 2417; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |

| | | | |
|----|-----|--|-----|
| QY | 1 | TCTAAGCGAAGATGTGCTGTTGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG | 60 |
| DB | 1 | TCTAAGCGAAGATGTGCTGTTGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG | 60 |
| QY | 61 | GAATTGGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG | 120 |
| DB | 61 | GAATTGGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG | 120 |
| QY | 121 | CCTAATCATCTTGGCCATCTCTGCTTAATTCGCTGGCTGGCTGGCTGGCTGGCTGGCT | 180 |
| DB | 121 | CCTAATCATCTTGGCCATCTCTGCTTAATTCGCTGGCTGGCTGGCTGGCTGGCTGGCT | 180 |
| QY | 181 | CGCGCTCATCGTTTCGAGTCGATTCATTCATGAGTTTGAATTCGCTGGCTGGCTGGCT | 240 |
| DB | 181 | CGCGCTCATCGTTTCGAGTCGATTCATTCATGAGTTTGAATTCGCTGGCTGGCTGGCT | 240 |
| QY | 241 | CACCGCCATCATGTTGCAAGATGTTGGTACAACTTCCTCAATCGTTTCGACAGCGCGC | 300 |
| DB | 241 | CACCGCCATCATGTTGCAAGATGTTGGTACAACTTCCTCAATCGTTTCGACAGCGCGC | 300 |
| QY | 301 | ATGGTATCGGTCGGCAGGATTTGGGCGGTACCGTCTATCCGCGGCTGATGATACGTC | 360 |
| DB | 301 | ATGGTATCGGTCGGCAGGATTTGGGCGGTACCGTCTATCCGCGGCTGATGATACGTC | 360 |
| QY | 361 | CGCGGGAATCATTTGGCTGCTGACGTTACTCAACATACCGGTCATATTCGTCGATCTG | 420 |
| DB | 361 | CGCGGGAATCATTTGGCTGCTGACGTTACTCAACATACCGGTCATATTCGTCGATCTG | 420 |
| QY | 421 | CGTGTCTCGCGCGCATCTTCAGTGGCCTGACCTCCATCTCCACCTACTCTGACACAA | 480 |

| | | | |
|----|------|--|------|
| Db | 421 | CGTGTCTCGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACTGCTGACCAA | 480 |
| QY | 481 | GGAGCTGTGGTTCGGGGGGCGGGCTCTTTCGCGCCAGCTTTCATCGCCATCGTGGCTGG | 540 |
| Db | 481 | GGAGCTGTGGTTCGGGGGGCGGGCTCTTTCGCGCCAGCTTTCATCGCCATCGTGGCTGG | 540 |
| QY | 541 | CTACATCAGTAGTTCGGTGGCTGGATCGTACGATACGAGGGCATTCGCATATTCGCGCT | 600 |
| Db | 541 | CTACATCAGTAGTTCGGTGGCTGGATCGTACGATACGAGGGCATTCGCATATTCGCGCT | 600 |
| QY | 601 | GGAGTCACTACTTCTGTGGTGGCTCAGTGAAGATCGATCGCTGTTTCTGGTGGCTGG | 660 |
| Db | 601 | GGAGTCACTACTTCTGTGGTGGCTCAGTGAAGATCGATCGCTGTTTCTGGTGGCTGG | 660 |
| QY | 661 | CGCAGCGCTTGTCTCTACTTCTACATGCTGTCGCGCTGGGCTGGCTGCTGCTGCTGCT | 720 |
| Db | 661 | CGCAGCGCTTGTCTCTACTTCTACATGCTGTCGCGCTGGGCTGGCTGCTGCTGCTGCT | 720 |
| QY | 721 | CAACCTGATACCCCTGCGAGCTTTCGTAATGCTGCTATTAAGGCGAGGTACTCGCGCGCT | 780 |
| Db | 721 | CAACCTGATACCCCTGCGAGCTTTCGTAATGCTGCTATTAAGGCGAGGTACTCGCGCGCT | 780 |
| QY | 781 | GCTGACAGCTACAGCACCTTCTACATCCTGGGACTGCTGCTTCTCCATGAGATCCCTT | 840 |
| Db | 781 | GCTGACAGCTACAGCACCTTCTACATCCTGGGACTGCTGCTTCTCCATGAGATCCCTT | 840 |
| QY | 841 | CGTGGGATTCACCGATACGACCAAGTGAACACATGCTGCTGGCTGGGAGTGTGCTGCT | 900 |
| Db | 841 | CGTGGGATTCACCGATACGACCAAGTGAACACATGCTGCTGGCTGGGAGTGTGCTGCT | 900 |
| QY | 901 | CCTTATGCGCTGGCGCACCTTCGCGCATTTGAGTCCGCTGCTGCGCGCAACAGATTCG | 960 |
| Db | 901 | CCTTATGCGCTGGCGCACCTTCGCGCATTTGAGTCCGCTGCTGCGCGCAACAGATTCG | 960 |
| QY | 961 | GAAGCTGTTTCTGCTGGCGGATTCGCTGGCGGATTCGCTGGCGGATTCGCTGGCGGAT | 1020 |
| Db | 961 | GAAGCTGTTTCTGCTGGCGGATTCGCTGGCGGATTCGCTGGCGGATTCGCTGGCGGAT | 1020 |
| QY | 1021 | GCTCACCATGCTGGCGGCTTGTGGCGGCTGGAGTGGAGCTTCTACTCGCTGTGGGATAC | 1080 |
| Db | 1021 | GCTCACCATGCTGGCGGCTTGTGGCGGCTGGAGTGGAGCTTCTACTCGCTGTGGGATAC | 1080 |
| QY | 1081 | TGCTACGCGAAGATCCACATTCCTCATTCGATCCGCTGCGAGCATCAGCCACAC | 1140 |
| Db | 1081 | TGCTACGCGAAGATCCACATTCCTCATTCGATCCGCTGCGAGCATCAGCCACAC | 1140 |
| QY | 1141 | TTGGTCTCTGTTCTTCTTGTGATCTGCATCTGCTGCTGGCGCTTCCAGTGGGAGTGTG | 1200 |
| Db | 1141 | TTGGTCTCTGTTCTTCTTGTGATCTGCATCTGCTGCTGGCGCTTCCAGTGGGAGTGTG | 1200 |
| QY | 1201 | GTAATGATCAAGCAGATCAAGCAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1260 |
| Db | 1201 | GTAATGATCAAGCAGATCAAGCAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1260 |
| QY | 1261 | GGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1320 |
| Db | 1261 | GGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1320 |
| QY | 1321 | CTGGCGGAGTGGCTTTCGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1380 |
| Db | 1321 | CTGGCGGAGTGGCTTTCGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1380 |
| QY | 1381 | GCAGATGGGCGCAGACCAAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG | 1440 |
| Db | 1381 | GCAGATGGGCGCAGACCAAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG | 1440 |
| QY | 1441 | GAAGAGAGCTGTGACGAGCTGGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG | 1500 |
| Db | 1441 | GAAGAGAGCTGTGACGAGCTGGCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG | 1500 |
| QY | 1501 | GCAGGATACCTGGCGCTGAGCTCCACCTGAGAGGATTTGTTATTTTGGCGCTTCTAAATGCT | 1560 |
| Db | 1501 | GCAGGATACCTGGCGCTGAGCTCCACCTGAGAGGATTTGTTATTTTGGCGCTTCTAAATGCT | 1560 |

QY 1561 GTTGATGATGTTGGTGTCTCCATGCAAGTGGGTGACCAAGTCTACTCCAGTCCCTC 1620
Db 1561 GTTGATGATGTTGGTGTCTCCATGCAAGTGGGTGACCAAGTCTACTCCAGTCCCTC 1620
QY 1621 CATGTCTTGGCTTTCCCAACAGTCAAGTGGATCCCGCAACATTTTAGAGCAATTCAG 1680
Db 1621 CATGTCTTGGCTTTCCCAACAGTCAAGTGGATCCCGCAACATTTTAGAGCAATTCAG 1680
QY 1681 AGAGCTTACTACTGGCTTTCCAGAACACTGCCGATGATGCTCGCGTTATGTTGGTG 1740
Db 1681 AGAGCTTACTACTGGCTTTCCAGAACACTGCCGATGATGCTCGCGTTATGTTGGTG 1740
QY 1741 CGATTACGGATACAGATAGCGGGAATGGCAACAGACGACGCTAGTGGATAATAATAC 1800
Db 1741 CGATTACGGATACAGATAGCGGGAATGGCAACAGACGACGCTAGTGGATAATAATAC 1800
QY 1801 GTGGAACAATAGTACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
Db 1801 GTGGAACAATAGTACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
QY 1861 CTACGAATATATGATCTCTTGGAGTGAATGCTTTGGTGTATCTTTGGCGGTGTAT 1920
Db 1861 CTACGAATATATGATCTCTTGGAGTGAATGCTTTGGTGTATCTTTGGCGGTGTAT 1920
QY 1921 CGGCTATTCTGGCGATGATATCAACAGTCTCTGCGTACGTTTGGTGTATCTTTGGCGGTGTAT 1980
Db 1921 CGGCTATTCTGGCGATGATATCAACAGTCTCTGCGTACGTTTGGTGTATCTTTGGCGGTGTAT 1980
QY 1981 GCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGA 2040
Db 1981 GCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGA 2040
QY 2041 TGCCGAAGGTCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGTCTACTACAGATT 2100
Db 2041 TGCCGAAGGTCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGTCTACTACAGATT 2100
QY 2101 CGGGAATTTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGACACAGTAAAGCGCT 2160
Db 2101 CGGGAATTTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGACACAGTAAAGCGCT 2160
QY 2161 CATTTGGGAATTAAGACTTGGATCTGACCTACCTGAGAGGCGCTTACACCAAGACATG 2220
Db 2161 CATTTGGGAATTAAGACTTGGATCTGACCTACCTGAGAGGCGCTTACACCAAGACATG 2220
QY 2221 GCTTGTTCGCATCTATAGGTGAAGAGCGCGATGAGTTCAATAGACCATCTAGAGAC 2280
Db 2221 GCTTGTTCGCATCTATAGGTGAAGAGCGCGATGAGTTCAATAGACCATCTAGAGAC 2280
QY 2281 CAAGGAGAACGATTCCTCCAGCAACTTCAATTCGAGAAAGAACTCTAAGCGTGGCAA 2340
Db 2281 CAAGGAGAACGATTCCTCCAGCAACTTCAATTCGAGAAAGAACTCTAAGCGTGGCAA 2340
QY 2341 GGGCTACATACGAACCGAGCGGTGTTGTTAAGGGAACGAACTTGAATTAACCCCA 2400
Db 2341 GGGCTACATACGAACCGAGCGGTGTTGTTAAGGGAACGAACTTGAATTAACCCCA 2400
QY 2401 AAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAA 2417

RESULT 2
AX799088
LOCUS AX799088 2417 bp DNA linear PAT 08-OCT-2003
DEFINITION Sequence 7 from Patent WO03054008.
ACCESSION AX799088
VERSION AX799088.1 GI:37605061
KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

REFERENCE
AUTHORS
TITLE

JOURNAL

FEATURES

source

Ephydroidea; Drosophilidae; Drosophila.

1. 2417

Perrault, C. and McBride, K.

Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy

Patent: WO 03054008-A 7 03-JUL-2003;

Comptigene Inc. (CA)

Location/Qualifiers

1. 2417

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/note="STT3, Acc# AF132552"

ORIGIN

Query Match 100.0%; Score 2417; DB 6; Length 2417;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTGCTGTGCAATTCAGATCGGTATATAATTTTCGAGTTACTGGCTG 60

Db 1 TCTAAGCGAAGATGTGCTGTGCAATTCAGATCGGTATATAATTTTCGAGTTACTGGCTG 60

QY 61 GAATTCGGACATGAATCGGACCGGAGATGCTGAACAGAAAGTGGTGGCTACAGCAG 120

Db 61 GAATTCGGACATGAATCGGACCGGAGATGCTGAACAGAAAGTGGTGGCTACAGCAG 120

QY 121 CTTAATCACCTTCGCGCATCTCTAATCGCTTGGTGGCGGATTTCTCTCGCTCTT 180

Db 121 CTTAATCACCTTCGCGCATCTCTAATCGCTTGGTGGCGGATTTCTCTCGCTCTT 180

QY 181 CGCGCTCATCCGTTTCGAGTGCATTCATCATGAGTTGATCGTGGTCACTACCGGCG 240

Db 181 CGCGCTCATCCGTTTCGAGTGCATTCATCATGAGTTGATCGTGGTCACTACCGGCG 240

QY 241 CACCGCTACATGTTGCGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 300

Db 241 CACCGCTACATGTTGCGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 300

QY 301 ATGGTATCCGCTCGGACGATTTGTTGGGGGTACCGCTCTATCCCGGCTGATGATTACGTC 360

Db 301 ATGGTATCCGCTCGGACGATTTGTTGGGGGTACCGCTCTATCCCGGCTGATGATTACGTC 360

QY 361 CGCGCGAATCCATTCGCTGCTGCACTACTCAATACCGTTCATATTCGAGCATCTG 420

Db 361 CGCGCGAATCCATTCGCTGCTGCACTACTCAATACCGTTCATATTCGAGCATCTG 420

QY 421 CGTGTTCCTGCGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

Db 421 CGTGTTCCTGCGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

QY 481 GGAGCTGTGTTTCGCGCGCGCTCTTCGCGCGAGTTCATTCGCGCATCTGCTGCGC 540

Db 481 GGAGCTGTGTTTCGCGCGCGCTCTTCGCGCGAGTTCATTCGCGCATCTGCTGCGC 540

QY 541 CTACATCAGTAGTTCGCTGCTGATCGTACGATTAAGAGGCGATTCGCAATTCGCGCT 600

Db 541 CTACATCAGTAGTTCGCTGCTGATCGTACGATTAAGAGGCGATTCGCAATTCGCGCT 600

QY 601 GCAGTTACCTACTTCTGTTGGTGGCTCAGTGAAGACTGATCCGCTGTTCTGCTGGC 660

Db 601 GCAGTTACCTACTTCTGTTGGTGGCTCAGTGAAGACTGATCCGCTGTTCTGCTGGC 660

QY 661 CGCAGCGCTTGTCTTCTACTTCTACATGTTGTCGCTGGGTGGGTGGGTGGGTGGGTGG 720

Db 661 CGCAGCGCTTGTCTTCTACTTCTACATGTTGTCGCTGGGTGGGTGGGTGGGTGGGTGG 720

QY 721 CAACCTGATACCTTCGCGCGCTCTTCGTAAGTCTCATTTATGGGAGGAGTCTCGCGCGCT 780

Db 721 CAACCTGATACCTTCGCGCGCTCTTCGTAAGTCTCATTTATGGGAGGAGTCTCGCGCGCT 780

QY 781 GCTGACGAGCTACAGCAGCTTCTACATCTGCTGGGAGTCTGCTGTTCTTCCATGAGATTC 840

Db 781 GCTGACGAGCTACAGCAGCTTCTACATCTGCTGGGAGTCTGCTGTTCTTCCATGAGATTC 840

Db 781 GCTGACAGCTACAGACCTTCTACATCCTGGAGCTGCTGTCTTCTCCATGCGAGATCCCGCTT 840
 Qy 841 CGTGGGATTCACACGATACGACAGATGAAACACATGCTGCTGGAGTGTGTGCT 900
 Db 841 CGTGGGATTCACACGATACGACAGATGAAACACATGCTGCTGGAGTGTGTGCT 900
 Qy 901 CTTATGCGCTGGGACCTTGGGCAATTTGAGTCCGCTGCTGCTGCGGCAACAGATTCG 960
 Db 901 CTTATGCGCTGGGACCTTGGGCAATTTGAGTCCGCTGCTGCTGCGGCAACAGATTCG 960
 Qy 961 GAAGCTGTTTCATGCTGGGAGATTTGCTGGGAGCTTGGGCTTCTTGTGGCGCTGCTG 1020
 Db 961 GAAGCTGTTTCATGCTGGGAGATTTGCTGGGAGCTTGGGCTTCTTGTGGCGCTGCTG 1020
 Qy 1021 GCTCACCATGCTGGGCTTGTGGGCTTGTGGGCTTGTGGGCTTGTGGGCTTGTGGGCTT 1080
 Db 1021 GCTCACCATGCTGGGCTTGTGGGCTTGTGGGCTTGTGGGCTTGTGGGCTTGTGGGCTT 1080
 Qy 1081 TGCTAGCGCAAGATCCATTCCTCATCTTGCATCCGTCGAGGATCAGCCACAC 1140
 Db 1081 TGCTAGCGCAAGATCCATTCCTCATCTTGCATCCGTCGAGGATCAGCCACAC 1140
 Qy 1141 TTGTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1200
 Db 1141 TTGTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1200
 Qy 1201 GTACTGATCAAGCAGATCAACAGAGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 1201 GTACTGATCAAGCAGATCAACAGAGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Qy 1261 GGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Db 1261 GGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Qy 1321 GCTGGCGGAGTGGCTTTTGGGAGCTTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 Db 1321 GCTGGCGGAGTGGCTTTTGGGAGCTTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 Qy 1381 GCGAATGGGACAGCAGATGACGACGACGACGACGACGACGACGACGACGACGACGAC 1440
 Db 1381 GCGAATGGGACAGCAGATGACGACGACGACGACGACGACGACGACGACGACGACGAC 1440
 Qy 1441 GAAGAAGACGCTGTACGACAGCTGCGAAGCTGGAAGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 1441 GAAGAAGACGCTGTACGACAGCTGCGAAGCTGGAAGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Qy 1501 GCAGGATCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 Db 1501 GCAGGATCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 Qy 1561 GTTCTGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 1561 GTTCTGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Qy 1621 CATTTGCTTGGCTTTCACACAGTCAAGATGATCCGCAACATTTTACAGATTTTAC 1680
 Db 1621 CATTTGCTTGGCTTTCACACAGTCAAGATGATCCGCAACATTTTACAGATTTTAC 1680
 Qy 1681 AGAGGCTTACTGCTGCTTTCGACAGACCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCT 1740
 Db 1681 AGAGGCTTACTGCTGCTTTCGACAGACCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCT 1740
 Qy 1741 GGTATGAGTACAGATGAGGAGATGGGCAACAGACAGCAGCTAGTGAATATATATAC 1800
 Db 1741 GGTATGAGTACAGATGAGGAGATGGGCAACAGACAGCAGCTAGTGAATATATATAC 1800
 Qy 1801 GTGGACATAGTACATAGCTGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 Db 1801 GTGGACATAGTACATAGCTGCTTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 Qy 1861 CTACGAATATGACATCTTTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 Db 1861 CTACGAATATGACATCTTTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920

Qy 1921 CGGCTATTTCTGGGATGATATCAACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 Db 1921 CGGCTATTTCTGGGATGATATCAACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 Qy 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTTACCGACCGCGTGAATTCAGGTTAGA 2040
 Db 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTTACCGACCGCGTGAATTCAGGTTAGA 2040
 Qy 2041 TGCGGAGGCTGCTCGGCGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
 Db 2041 TGCGGAGGCTGCTCGGCGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
 Qy 2101 CGGGAATTAAGCTTGGACTACAGAGTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
 Db 2101 CGGGAATTAAGCTTGGACTACAGAGTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
 Qy 2161 CATTTGGAATAAGGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Db 2161 CATTTGGAATAAGGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Qy 2221 GCTTGTTCGCTATAGGCTGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
 Db 2221 GCTTGTTCGCTATAGGCTGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
 Qy 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
 Db 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
 Qy 2341 GGGCTACATAGCAACCGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
 Db 2341 GGGCTACATAGCAACCGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
 Qy 2401 AAAAAAAAAAAAAAAAAA 2417
 Db 2401 AAAAAAAAAAAAAAAAAA 2417

RESULT 3
 CO589353
 LOCUS Sequence 17111 from Patent WO0171042. linear PAT 02-FEB-2004
 DEFINITION CO589353
 ACCESSION CO589353
 VERSION CO589353.1 GI:41648215
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1
 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 Detection kits, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 Patent: WO 0171042-A 17111 27-SEP-2001;
 PE Corporation (NY) (US)
 JOURNAL location/Qualifiers
 FEATURES
 source
 1. 2699
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"

ORIGIN
 Query Match 99.7%; Score 2409; DB 6; Length 2699;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TCTAGCGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 Db 10 TCTAGCGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 69
 Qy 61 GAATTTGGACATGATCGAGCGGAGATGCTGCAACAGCAAGTGGCTGCTGCTGCTGCTGCTGCT 120

Db 70 GAATTGGGACATGAATCGGACGCGGAGAGATGCTGAACAGCAAGGTGGCTGCTACAGCAG 129
 QY 121 CTTAATCACTTCGCGCATCTCTGTAATCGCTGCGTGGCGGATTTTCTCTCTCGCCCTTT 180
 Db 130 CTTAATCACTTCGCGCATCTCTGTAATCGCTGCGTGGCGGATTTTCTCTCTCGCCCTTT 189
 QY 181 CGCGCTCATCGTTTCGAGTGAATTCATCATGAGTTTGAATCCGTTGGTTCACTACCGGGC 240
 Db 190 CGCGCTCATCGTTTCGAGTGAATTCATCATGAGTTTGAATCCGTTGGTTCACTACCGGGC 249
 QY 241 CACCGCTACATGTTGCGAGATGTTGTTGTAACAATCTTCTCAACTGGTTTCAGACAGCGCGC 300
 Db 250 CACCGCTACATGTTGCGAGATGTTGTTGTAACAATCTTCTCAACTGGTTTCAGACAGCGCGC 309
 QY 301 ATGGTATCGCTTCGCGAGATGTTGCGGCTGACGCTCTATCCCGGCTGATGATATGCTC 360
 Db 310 ATGGTATCGCTTCGCGAGATGTTGCGGCTGACGCTCTATCCCGGCTGATGATATGCTC 369
 QY 361 CGCGGGAATCCATTTGGCTGCTGACGTAATCAACATACCGGCTCATATTCGTTGACATCTG 420
 Db 370 CGCGGGAATCCATTTGGCTGCTGACGTAATCAACATACCGGCTCATATTCGTTGACATCTG 429
 QY 421 CGTGTCTCGGCGCGATCTTCAGTGGCTGACCTTCATCTCCACCTACCTGCTGACCAA 480
 Db 430 CGTGTCTCGGCGCGATCTTCAGTGGCTGACCTTCATCTCCACCTACCTGCTGACCAA 489
 QY 481 GGAGCTGTGGTCCGCGGCGCGGCTCTTCCGCGCGAGCTTTCATCGCCATCGTGGCTGG 540
 Db 490 GGAGCTGTGGTCCGCGGCGCGGCTCTTCCGCGCGAGCTTTCATCGCCATCGTGGCTGG 549
 QY 541 CTACATCAGTAGTGGTGGCTGGATCGTACGATTAACAGGCGATTCGCTATTCGCGCT 600
 Db 550 CTACATCAGTAGTGGTGGCTGGATCGTACGATTAACAGGCGATTCGCTATTCGCGCT 609
 QY 601 GCAGTTCACCTACTTCTGTGGGTGGCTCAGTGAAGACTGGATCCGTTCTGTGGTGGC 660
 Db 610 GCAGTTCACCTACTTCTGTGGGTGGCTCAGTGAAGACTGGATCCGTTCTGTGGTGGC 669
 QY 661 CGCAGCGCTTTGCTCTACTTCTACATGGTGTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 720
 Db 670 CGCAGCGCTTTGCTCTACTTCTACATGGTGTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 729
 QY 721 CAACCTGATACCCCTGACGCTTCTGTAATGCTGCTGATGAGGAGTACTCGCGCGGTCT 780
 Db 730 CAACCTGATACCCCTGACGCTTCTGTAATGCTGCTGATGAGGAGTACTCGCGCGGTCT 789
 QY 781 GCTGACGAGTACAGACCTTCTACATCTGGGACTGCTGTTCTGATGAGATTCGCTT 840
 Db 790 GCTGACGAGTACAGACCTTCTACATCTGGGACTGCTGTTCTGATGAGATTCGCTT 849
 QY 841 CGTGGGATTCACCGGATGACGACGAGTGAACATGCTGCGCTGGGAGTGTGTTGCT 900
 Db 850 CGTGGGATTCACCGGATGACGACGAGTGAACATGCTGCGCTGGGAGTGTGTTGCT 909
 QY 901 CTTATGCGCTGGCGACCTTGGCGCATTTGAGTCCGTTGCTGCTGGCGAACGAGTTCCG 960
 Db 910 CTTATGCGCTGGCGACCTTGGCGCATTTGAGTCCGTTGCTGCTGGCGAACGAGTTCCG 969
 QY 961 GAAGCTGTTCACTGCTGGCGGATTTGCTGGTGGGCTTGGGCTCTTGTGGCGCTGGGCT 1020
 Db 970 GAAGCTGTTCACTGCTGGCGGATTTGCTGGTGGGCTTGGGCTCTTGTGGCGCTGGGCT 1029
 QY 1021 GCTCAACATGCTGGCGGTTGGGCGCTGGGAGTGGAGCTTCTGCTGCTGGGATC 1080
 Db 1030 GCTCAACATGCTGGCGGTTGGGCGCTGGGAGTGGAGCTTCTGCTGCTGGGATC 1089
 QY 1081 TGGCTACGCCAAGATCCATTCCTCCATCATTCGCTGCGGAGCATCAGCCACAC 1140
 Db 1090 TGGCTACGCCAAGATCCATTCCTCCATCATTCGCTGCGGAGCATCAGCCACAC 1149
 QY 1141 TTGGTTCCTGTTCTTCTTTGATCTGCAATCTGGTGTGGCGCTTCCAGTGGAGTGTG 1200
 Db 1150 TTGGTTCCTGTTCTTCTTTGATCTGCAATCTGGTGTGGCGCTTCCAGTGGAGTGTG 1209

QY 1201 GTACTGCATCAAGCAGATCAACGACGAGCGGTTTTCTGGTGTCTGTACGCCATCAGTGC 1260
 Db 1210 GTACTGCATCAAGCAGATCAACGACGAGCGGTTTTCTGGTGTCTGTACGCCATCAGTGC 1269
 QY 1261 GGTTTACTCTCGCTGGTGTGATGCGGTTTGTGATGTTGACCTCAGCCGCGTGTGTCAT 1320
 Db 1270 GGTTTACTCTCGCTGGTGTGATGCGGTTTGTGATGTTGACCTCAGCCGCGTGTGTCAT 1329
 QY 1321 GCTGGCGGAGTGGCTTTTTCGGACTGTTGGATGTTTCTCTCAAGAGGATTCGTCTAA 1380
 Db 1330 GCTGGCGGAGTGGCTTTTTCGGACTGTTGGATGTTTCTCTCAAGAGGATTCGTCTAA 1389
 QY 1381 GCGAATGGGACACGCCATNAGCCGACGACGAGTGGATGAGCTGAGGATTCGATTTGA 1440
 Db 1390 GCGAATGGGACACGCCATNAGCCGACGACGAGTGGATGAGCTGAGGATTCGATTTGA 1449
 QY 1441 GAAGAAGACGCTGTACGACCAAGGCTGGCAAGCTGAAGCATCTGTAAGCATGATGCCCA 1500
 Db 1450 GAAGAAGACGCTGTACGACCAAGGCTGGCAAGCTGAAGCATCTGTAAGCATGATGCCCA 1509
 QY 1501 GCAGGATACGCGCTGAGCTCCACCTGAAGATGATTTGTTATTTGGCGGTTCTAATGCT 1560
 Db 1510 GCAGGATACGCGCTGAGCTCCACCTGAAGATGATTTGTTATTTGGCGGTTCTAATGCT 1569
 QY 1561 GTTGATGATGTTCTGCTGTCCACTGTCAGCTGGGTGACGACCAATGCTTCTCAGTCCCTC 1620
 Db 1570 GTTGATGATGTTCTGCTGTCCACTGTCAGCTGGGTGACGACCAATGCTTCTCAGTCCCTC 1629
 QY 1621 CATTTCTTGGCTTTCACACAGTGAAGTGGATCCGCAACATTTTAGACGATTTAG 1680
 Db 1630 CATTTCTTGGCTTTCACACAGTGAAGTGGATCCGCAACATTTTAGACGATTTAG 1689
 QY 1681 AGAGCTTACTGCTGCTTTCGCAAGACACTGCGGATGATGCTCGGCTTATGCTTGGTG 1740
 Db 1690 AGAGCTTACTGCTGCTTTCGCAAGACACTGCGGATGATGCTCGGCTTATGCTTGGTG 1749
 QY 1741 GGAATACGATACGAGATAGCGGAAATGGCAACAGCAACGAGCTAGTGAATTAATAC 1800
 Db 1750 GGAATACGATACGAGATAGCGGAAATGGCAACAGCAACGAGCTAGTGAATTAATAC 1809
 QY 1801 GTGGAACAAATAGTCACATAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
 Db 1810 GTGGAACAAATAGTCACATAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1869
 QY 1861 CTACGAATTAATGATCATCTTTCAGCTGGATGATGTTTGGTGTCTTTGGCGGCTGAT 1920
 Db 1870 CTACGAATTAATGATCATCTTTCAGCTGGATGATGTTTGGTGTCTTTGGCGGCTGAT 1929
 QY 1921 CGGCTATTCTGGCGATGATATCAACAAAGTTCCTGTGGATGCTCCGAATTCCTGAGGAGA 1980
 Db 1930 CGGCTATTCTGGCGATGATATCAACAAAGTTCCTGTGGATGCTCCGAATTCCTGAGGAGA 1989
 QY 1981 GCATCCGAGACATTAAGGAAGCGATTTACCGACGCGGTGAATTCAGGCTAGA 2040
 Db 1990 GCATCCGAGACATTAAGGAAGCGATTTACCGACGCGGTGAATTCAGGCTAGA 2049
 QY 2041 TCCCGAAGTCTCGCGGCTTCTCAACTGCTTATGTAACAATTAAGCTACTACAGATT 2100
 Db 2050 TCCCGAAGTCTCGCGGCTTCTCAACTGCTTATGTAACAATTAAGCTACTACAGATT 2109
 QY 2101 CGGGGAATTGAAGTTGACTACAGAGTCCATCTGATATGATGCGACACAGTAACCGCT 2160
 Db 2110 CGGGGAATTGAAGTTGACTACAGAGTCCATCTGATATGATGCGACACAGTAACCGCT 2169
 QY 2161 CATTTGGGAATTAAGGATTCGATCTGACCTACCTGGAGAGGCGCTCACCAAGAACCTG 2220
 Db 2170 CATTTGGGAATTAAGGATTCGATCTGACCTACCTGGAGAGGCGCTCACCAAGAACCTG 2229
 QY 2221 GCTTGTTCGCTATTAAGGTTGAAGGCGGATGAGTTCAATAGACCATCTGAAGAC 2280
 Db 2230 GCTTGTTCGCTATTAAGGTTGAAGGCGGATGAGTTCAATAGACCATCTGAAGAC 2289

QY 2281 CAAGGAGAGAACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTTAAGCGTCGCAA 2340
Db 2290 CAAGGAGAGAACGATTCCTCCAGCAAACTTCAATTCGAGAAAGAACTTAAGCGTCGCAA 2349
QY 2341 GGGCTACATACGAAACCGAGCGGTGTGTAGGGAACGAAACGAACTTCAAAATAACCCA 2400
Db 2350 GGGCTACATACGAAACCGAGCGGTGTGTAGGGAACGAAACGAACTTCAAAATAACCCA 2409
QY 2401 AAAAAAAA 2409
Db 2410 AAAAAAAA 2418

RESULT 4
CQ589352/c
LOCUS
DEFINITION
SEQUENCE 17110 from Patent WO0171042.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
PE Corporation (NY) (US)
Location/Qualifiers
1. 4922
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN
Query Match 60.5%; Score 1463; DB 6; Length 4922;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAAGCGAAGATGTCGTGTCATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 60
Db 3913 TCTAAGCGAAGATGTCGTGTCATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 3854
QY 61 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGGTGGCTACAGCAG 120
Db 3853 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGGTGGCTACAGCAG 3794
QY 121 CCTAATCACCTTCGCCATCTGCTATTCGCTGCGCTGGCGGATTTTCTCTCGCTCTT 180
Db 3793 CCTAATCACCTTCGCCATCTGCTATTCGCTGCGCTGGCGGATTTTCTCTCGCTCTT 3734
QY 181 CGCGCTCATCGCTTCGAGTCGATATTCGAGTTGATTCGCTGCTCAATACCGGCT 240
Db 3733 CGCGCTCATCGCTTCGAGTCGATATTCGAGTTGATTCGCTGCTCAATACCGGCT 3674
QY 241 CACGCGCTACATGCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Db 3673 CACGCGCTACATGCTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3614
QY 301 ATGTATTCGCTCGGCGAGATTTGGCGGCTACCGCTATTCGCGGCTGATATACGTC 360
Db 3613 ATGTATTCGCTCGGCGAGATTTGGCGGCTACCGCTATTCGCGGCTGATATACGTC 3554
QY 361 CGCGCGAATCCATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 3553 CGCGCGAATCCATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3494
QY 421 CGTGTTCCTGGCGCGATCTTTCAGTGGCTGACCTCCATCTCCAGCTCTGCTGACCAA 480
Db 3493 CGTGTTCCTGGCGCGATCTTTCAGTGGCTGACCTCCATCTCCAGCTCTGCTGACCAA 3434

QY 481 GGAGCTGTGCTCGCGGGCGCGCGCTCTTTCGCGCGAGCTTTCATCGCCATTCGTCCTGG 540
Db 3433 GGAGCTGTGCTCGCGGGCGCGCGCTCTTTCGCGCGAGCTTTCATCGCCATTCGTCCTGG 3374
QY 541 CTACATCAGTAGGTGCTGGCTGGATCGTACGATACGAGGCGATTCGCATATTCGCGCT 600
Db 3373 CTACATCAGTAGGTGCTGGCTGGATCGTACGATACGAGGCGATTCGCATATTCGCGCT 3314
QY 601 GCAGTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 3313 GCAGTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3254
QY 661 CGAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 3253 CGAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3194
QY 721 CAACCTGATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 3193 CAACCTGATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3134
QY 781 GCTGACCTGATACGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 3133 GCTGACCTGATACGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3074
QY 841 CGTGGGATTCGAACCGATACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 3073 CGTGGGATTCGAACCGATACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3014
QY 901 CTTATGCGCTGCGGACCTTTCGCGCATTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 3013 CTTATGCGCTGCGGACCTTTCGCGCATTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2954
QY 961 GAAGCTGTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 2953 GAAGCTGTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2894
QY 1021 GCTCACCATGCTGGGGCTTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 2893 GCTCACCATGCTGGGGCTTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2834
QY 1081 TGCTACGCAAGATCCCAATTCCTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 2833 TGCTACGCAAGATCCCAATTCCTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2774
QY 1141 TTGGTTCTGCTTCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 2773 TTGGTTCTGCTTCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2714
QY 1201 GTACTGCATCAAGAGATCAACGAGCGGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 2713 GTACTGCATCAAGAGATCAACGAGCGGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2654
QY 1261 GGTTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 2653 GGTTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2594
QY 1321 GCTGGCGGAGTGGGCTTTCGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 2593 GCTGGCGGAGTGGGCTTTCGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2534
QY 1381 GCGAATGGCGACAGCGATACGCGGAGCGGACCGAGAGTGAAGTGAAGTGAAGTGAAGTGA 1440
Db 2533 GCGAATGGCGACAGCGATACGCGGAGCGGACCGAGAGTGAAGTGAAGTGAAGTGAAGTGA 2474
QY 1441 GAAGAAGAGCTGTACGCAAGG 1463
Db 2473 GAAGAAGAGCTGTACGCAAGG 2451

RESULT 5
AC007853/c
LOCUS
162921 bp DNA linear INV 27-FEB-2001

| | |
|------------|--|
| DEFINITION | Drosophila melanogaster, chromosome 3R, region 96B-96C, BAC clone BAC03J02, complete sequence. |
| ACCESSION | AC007853 |
| VERSION | AC007853.4 |
| KEYWORDS | HTG |
| SOURCE | GI:13129410 |
| ORGANISM | Drosophila melanogaster (fruit fly) |
| REFERENCE | Drosophila melanogaster |
| AUTHORS | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. |
| REFERENCE | 1 (bases 1 to 162921) |
| AUTHORS | Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsset, V., Doup, L.S., Doyle, C., Dresnek, D., Farfan, D., Ferrier, S., Friese, E., Galle, R.F., Garg, N.S., George, R.A., Gonzales, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacle, J., Patagas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C. |
| TITLE | Sequencing of Drosophila chromosome 3R, region 96B-96C |
| JOURNAL | Unpublished |
| REFERENCE | 2 (bases 1 to 162921) |
| AUTHORS | Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Butenhoff, C., Champs, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA |
| COMMENT | On Feb 27, 2001 this sequence version replaced gi:5670481. Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu . Location/Qualifiers 1. 162921 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /strain="y; cn bw sp" /db_xref="taxon:7227" /chromosomes="3R" /map="96B-96C" /clone="BAC03J02 (D766)" /clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACs.6)" |
| FEATURES | source |
| source | Query Match 60.5%; Score 1463; DB 3; Length 162921; Best Local Similarity 100.0%; Pred. No. 0; Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| ORIGIN | |
| Qy | 1 TCTACGAGAGATGTCGTTGCAATTCAGATCGTTATATATTTTCAGTACTGCGCTG 60 |
| Db | 108551 TCTAACGAGAGATGTCGTTGCAATTCAGATCGTTATATATTTTCAGTACTGCGCTG 108492 |
| Qy | 61 GAATTGGGACATGAATCGGACCGGAGATGCTGAACAGCAAGGTGGCTGGCTACAGCAG 1200 |
| Db | 108491 GAATTGGGACATGAATCGGACCGGAGATGCTGAACAGCAAGGTGGCTGGCTACAGCAG 108432 |
| Qy | 121 CTAATACCTTCGCGCATCTCTGTAATCGCTGGCTGGCGGATTTTCTCTCGCTCTT 180 |
| Db | 108431 CTAATACCTTCGCGCATCTCTGTAATCGCTGGCTGGCGGATTTTCTCTCGCTCTT 108372 |
| Qy | 181 CGCGCTCATCGCTTCGAGTCGATATTCATGAGTTGATCCGTTGGTTCAATACCGGGC 240 |
| Db | 108371 CGCGCTCATCGCTTCGAGTCGATATTCATGAGTTGATCCGTTGGTTCAATACCGGGC 108312 |
| Qy | 241 CACCGCTACATGCTGGCAGATGGTTGTGTAACAATCTCTCACTGGTTCCAGCAGCGCGC 300 |
| Db | 108311 CACCGCTACATGCTGGCAGATGGTTGTGTAACAATCTCTCACTGGTTCCAGCAGCGCGC 108252 |
| Qy | 301 ATGGTATCGCTCGGACAGATGTTGGCGGTACCGTCTATCCCGGCTGATGATAGTC 360 |
| Db | 108251 ATGGTATCGCTCGGACAGATGTTGGCGGTACCGTCTATCCCGGCTGATGATAGTC 108192 |
| Qy | 361 CGCGCGAATCCATTTGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG 420 |
| Db | 108191 CGCGCGAATCCATTTGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG 108132 |
| Qy | 421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTGCATCTCCACCTACCTGCTGACAA 480 |
| Db | 108131 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTGCATCTCCACCTACCTGCTGACAA 108072 |
| Qy | 481 GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCATCGGCTGG 540 |
| Db | 108071 GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCATCGGCTGG 108012 |
| Qy | 541 CTACATAGTAGTTCGTTGGTGGATCGTACGATAAGAGGGCATTCGCCATATTCGCGCT 600 |
| Db | 108011 CTACATAGTAGTTCGTTGGTGGATCGTACGATAAGAGGGCATTCGCCATATTCGCGCT 107952 |
| Qy | 601 CGAGTTCACTTCTTCCTGCTGGTGGCTCAGTCAAGACTGGATCCGTTCTGCTCGGC 660 |
| Db | 107951 CGAGTTCACTTCTTCCTGCTGGTGGCTCAGTCAAGACTGGATCCGTTCTGCTCGGC 107892 |
| Qy | 661 CGCAGCGCTTTCCTACTTCTAATAGTTCGCGCTCGCGGTGGCTACGTTTCATCAT 720 |
| Db | 107891 CGCAGCGCTTTCCTACTTCTAATAGTTCGCGCTCGCGGTGGCTACGTTTCATCAT 107832 |
| Qy | 721 CAACCTGATACCCCTGCGACGTTCTGCTACTCTCATATGGGAGGTACTCGCGCTCT 780 |
| Db | 107831 CAACCTGATACCCCTGCGACGTTCTGCTACTCTCATATGGGAGGTACTCGCGCTCT 107772 |
| Qy | 781 GCTGACCAAGCTACAGCACTTCTACATCTCGGACTGCTGTTCTCCATGAGATTCGCTT 840 |
| Db | 107771 GCTGACCAAGCTACAGCACTTCTACATCTCGGACTGCTGTTCTCCATGAGATTCGCTT 107712 |
| Qy | 841 CGTGGGATTCGAACCGATACCGACAGTGAACATCGCTGCGTGGGAGTGTGTTGCT 900 |
| Db | 107711 CGTGGGATTCGAACCGATACCGACAGTGAACATCGCTGCGTGGGAGTGTGTTGCT 107652 |
| Qy | 901 CTTATGCGCGTGGCGACCTTTCGCGCATTTGCGCTGCTGCTGCGCAACGAGTTCCG 960 |
| Db | 107651 CTTATGCGCGTGGCGACCTTTCGCGCATTTGCGCTGCTGCTGCGCAACGAGTTCCG 107592 |
| Qy | 961 GAAGCTGTTCATCGTCGCGGATTTGCTGTTGGGGTGGGCTCTTCTGCGCGCTCGTGT 1020 |
| Db | 107591 GAAGCTGTTCATCGTCGCGGATTTGCTGTTGGGGTGGGCTCTTCTGCGCGCTCGTGT 107532 |
| Qy | 1021 GCTACCATGCTGGCGTGTGGCGCGGTGGAGTGGACGCTTCTACTCGTGTGGGATAC 1080 |
| Db | 107531 GCTACCATGCTGGCGTGTGGCGCGGTGGAGTGGACGCTTCTACTCGTGTGGGATAC 107472 |
| Qy | 1081 TGGCTACCGCAGATCCACATTCCTCCATTCGATCGTGGAGCATCAGCCACAC 1140 |
| Db | 107471 TGGCTACCGCAGATCCACATTCCTCCATTCGATCGTGGAGCATCAGCCACAC 107412 |
| Qy | 1141 TTGGTCTCTGTTCTTTTGTGATCTGCACATCTCTGTTGGCGCTTCCAGTGGAGTGTG 1200 |

Db 107411 TTGGTCTCTGTTCTTCTTGTGATCTGCATCTCTGTTGGCTTCCAGTGGAGTGTG 107352
 Qy 1201 GTACTGCATCAAGCAGATCAACGACGAGCGGTTTCTGTTGCTGCTGTACGCCATCAGTGC 1260
 Db 107351 GTACTGCATCAAGCAGATCAACGACGAGCGGTTTCTGTTGCTGCTGTACGCCATCAGTGC 107292
 Qy 1261 GGTTTACTCTGCTGGTGTGATGAGTGCGTTTGTGATGACCTCACCGCGTGGTGTGAT 1320
 Db 107291 GGTTTACTCTGCTGGTGTGATGAGTGCGTTTGTGATGACCTCACCGCGTGGTGTGAT 107232
 Qy 1321 GCTGGCCGAGTGGCTTTTCGGGACGTTTGGGATGTTTCTCAGAGAGATTTCGTCTAA 1380
 Db 107231 GCTGGCCGAGTGGCTTTTCGGGACGTTTGGGATGTTTCTCAGAGAGATTTCGTCTAA 107172
 Qy 1381 CGGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATGAAGCTGAGATTCCATTGA 1440
 Db 107171 CGGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATGAAGCTGAGATTCCATTGA 107112
 Qy 1441 GAAGAAGACGCTGTACGACAAAGG 1463
 Db 107111 GAAGAAGACGCTGTACGACAAAGG 107089

RESULT 6

AC008206/6
 LOCUS AC008206 181132 bp DNA linear INV 24-FEB-2001
 DEFINITION Drosophila melanogaster, chromosome 3R, region 96B-96B, BAC clone
 AC008206
 ACCESSION AC008206
 VERSION AC008206.10 GI:13122717
 KEYWORDS HTG.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 181132)
 AUTHORS Celisneris,E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Bueam,D.A., Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Honck,J., Hoekins,R.A., Hostin,D., Howland,T.J., Ibegwan,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacle,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
 Sequencing of Drosophila chromosome 3R, region 96B-96B
 Unpublished

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 181132)
 Celisneris,E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Bubenhoef,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoekins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacle,J.M., Park,S., Pfeiffer,B., Poon,L., Sequelira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
 Direct Submission
 Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 24, 2001 this sequence version replaced gi:7208834.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT
 This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bagp@fruitfly.berkeley.edu.
 FEATURES
 source 1. 181132
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="96B-96B"
 /clone="BACR03115 (D765)"
 /clone_lib="RPCL-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"

ORIGIN

Query Match 60.5%; Score 1463; DB 3; Length 181132;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TCTAAGCGAAGATGTGCTTGCATTTTCAGATCGGTTATTAATTTTCAGTTACTGGCTG 60
 Db 171883 TCTAAGCGAAGATGTGCTTGCATTTTCAGATCGGTTATTAATTTTCAGTTACTGGCTG 171824
 Qy 61 GAATTCGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
 Db 171823 GAATTCGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 171764
 Qy 121 CTTAATCACTTCGCGCATCTCTCTAATCGCTGGCTGGCGGATTTCTCTCGCCTCTT 180
 Db 171763 CTTAATCACTTCGCGCATCTCTCTAATCGCTGGCTGGCGGATTTCTCTCGCCTCTT 171704
 Qy 181 CGCGCTCATCGCTTCGAGTCGATTTATCATGAGTTTCATCGCTGGTTCAACTACCGGGC 240
 Db 171703 CGCGCTCATCGCTTCGAGTCGATTTATCATGAGTTTCATCGCTGGTTCAACTACCGGGC 171644
 Qy 241 CACCGCTACATGGTGGCGAAGTGGTTGGTACAACTTCCTCAACTGGTTGGACGAGCGCGC 300
 Db 171643 CACCGCTACATGGTGGCGAAGTGGTTGGTACAACTTCCTCAACTGGTTGGACGAGCGCGC 171584
 Qy 301 ATGCTATCGCTCGCGGAGGATTTGGGGGATCGCTCTATCCCGGCTGATGATTAGTCTC 360
 Db 171583 ATGCTATCGCTCGCGGAGGATTTGGGGGATCGCTCTATCCCGGCTGATGATTAGTCTC 171524
 Qy 361 CGCGGAATTCATTTGGCTGCTGACGATCACTAATACCGGTCCATATTCGTGACATCTG 420
 Db 171523 CGCGGAATTCATTTGGCTGCTGACGATCACTAATACCGGTCCATATTCGTGACATCTG 171464
 Qy 421 CGGTTTCCTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
 Db 171463 CGGTTTCCTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 171404
 Qy 481 GGAGCTGTGGTTCGCGGGGCGCGGCTCTTTCGCGCGGAGCTTCATTCGCCATCTGGCTGG 540
 Db 171403 GGAGCTGTGGTTCGCGGGGCGCGGCTCTTTCGCGCGGAGCTTCATTCGCCATCTGGCTGG 171344
 Qy 541 CTACATCAGTAGTGGTGGCTGGATCTGTAGATAACGAGGGCATTTGCCATATTTCGCCCT 600
 Db 171343 CTACATCAGTAGTGGTGGCTGGATCTGTAGATAACGAGGGCATTTGCCATATTTCGCCCT 171284
 Qy 601 CGAGTTCACTACTTCTCTGTTGGTGGCTCAGTGAAGACTGGATTCGGTTCGGTGGCG 660
 Db 171283 CGAGTTCACTACTTCTCTGTTGGTGGCTCAGTGAAGACTGGATTCGGTTCGGTGGCG 171224
 Qy 661 CGCAGCGGCTTTTGTCTTCTTCTTCTACATGTTGTTCGCCCTGGGGTGGCTACGTTTCATCAT 720
 Db 171223 CGCAGCGGCTTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 171164
 Qy 721 CAACCTGATACCGCTGCAGCTTCTGTAAGTCTGCTATTAATGGGAGGACTTCGCCGCTCT 780
 Db 171163 CAACCTGATACCGCTGCAGCTTCTGTAAGTCTGCTATTAATGGGAGGACTTCGCCGCTCT 171104

1..2953
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

Query Match 28.8%; Score 696; DB 6; Length 2953;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 696; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

QY 1 TCTAAGCGAAGATGTCTGTTGCAATTCAGATCGGTTATATATTTTCAGTTACTTGGCTG 60
DB 696 TCTAAGCGAAGATGTCTGTTGCAATTCAGATCGGTTATATATTTTCAGTTACTTGGCTG 637
QY 61 GAATTCGGACATGAATTCGAGCGCGAAGATGCTGAACAGCAAGGTGGTGGCTACAGAG 120
DB 636 GAATTCGGACATGAATTCGAGCGCGAAGATGCTGAACAGCAAGGTGGTGGCTACAGAG 577
QY 121 CCTAATACCTTCGCCATCTCTGTAATCGCTGGTGGCGGATTTCTCTCGCTCTTT 180
DB 576 CCTAATACCTTCGCCATCTCTGTAATCGCTGGTGGCGGATTTCTCTCGCTCTTT 517
QY 181 CGCGTCATCCGTTTCGATTCGATTCATGAGTTTTCATCGGTTTCAACTACCGGCG 240
DB 516 CGCGTCATCCGTTTCGATTCGATTCATGAGTTTTCATCGGTTTCAACTACCGGCG 457
QY 241 CACCGCTACATGCTGAGATGTTGGTACAACTTCCTCAACTGGTTTCGACGAGCGCG 300
DB 456 CACCGCTACATGCTGAGATGTTGGTACAACTTCCTCAACTGGTTTCGACGAGCGCG 397
QY 301 ATGGTATCCGTCGCGGAGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTCGTC 360
DB 396 ATGGTATCCGTCGCGGAGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTCGTC 337
QY 361 CGCGGCAATTCATTTGCTGCTGACGATCAACATACCGGTCCATATTCGTGACATCTG 420
DB 336 CGCGGCAATTCATTTGCTGCTGACGATCAACATACCGGTCCATATTCGTGACATCTG 277
QY 421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACAA 480
DB 276 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACAA 217
QY 481 GGAGCTGTGTTCGCGGCGCGGCTCTTCGCGCGGAGCTTCATCGCATCTGCTGCTG 540
DB 216 GGAGCTGTGTTCGCGGCGCGGCTCTTCGCGCGGAGCTTCATCGCATCTGCTGCTG 157
QY 541 CTACATCAGTAGTTCGGTGGCTGATTCGTAAGATTCGTAAGATTCGTAAGATTCGTAAG 600
DB 156 CTACATCAGTAGTTCGGTGGCTGATTCGTAAGATTCGTAAGATTCGTAAGATTCGTAAG 97
QY 601 GCAGTTCACTACTCTCTGTTGCTCTTCTACATGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 96 GCAGTTCACTACTCTCTGTTGCTCTTCTACATGCTGCTGCTGCTGCTGCTGCTGCTG 37
QY 661 CGCAGCGCTTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 696
DB 36 CGCAGCGCTTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1

RESULT 10
LOCUS CQ596795/c 2785 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 24553 from Patent WO0171042.
ACCESSION CQ596795
VERSION CQ596795.1 GI:41653176
KEYWORDS
SOURCE Drosophila sp.
ORGANISM Drosophila sp.
REFERENCE Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
AUTHORS Detection kits, such as nucleic acid arrays, for detecting the
TITLE expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 017042-A 24577 27-SEP-2001;
FEATURES PE Corporation (NY) (US)
Location/Qualifiers

RESULT 9
LOCUS CQ596819/c 2953 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 24577 from Patent WO0171042.
ACCESSION CQ596819
VERSION CQ596819.1 GI:41653192
KEYWORDS
SOURCE Drosophila sp.
ORGANISM Drosophila sp.
REFERENCE Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
AUTHORS Detection kits, such as nucleic acid arrays, for detecting the
TITLE expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 017042-A 24577 27-SEP-2001;
FEATURES PE Corporation (NY) (US)
Location/Qualifiers

TITLE Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more *Drosophila* genes and uses thereof
JOURNAL Patent: WO 01/1042-A 24553 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES Location/Qualifiers
source 1..2785
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN

Query Match 10.3%; Score 249; DB 6; Length 2785;
Best Local Similarity 100.0%; Pred. No. 5.3e-136;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1876 ATCTTCAACGAGTTCCTGTGGATGCTCCGAATTCCTGAGGAGGATCCCAAGGACAT 1995
Db |||||||
2795 ATCTTCAACGAGTTCCTGTGGATGCTCCGAATTCCTGAGGAGGATCCCAAGGACAT 2726
QY 1936 TGATATCAACGAGTTCCTGTGGATGCTCCGAATTCCTGAGGAGGATCCCAAGGACAT 1995
Db |||||||
2725 TGATATCAACGAGTTCCTGTGGATGCTCCGAATTCCTGAGGAGGATCCCAAGGACAT 2666
QY 1996 TAAGAAAGCGATTACTTACCGACCGCGGTGAATTCAGGATAGATCCCGAAGTGTCTCC 2055
Db |||||||
2665 TAAGAAAGCGATTACTTACCGACCGCGGTGAATTCAGGATAGATCCCGAAGTGTCTCC 2606
QY 2056 GGCCTGTCTCAACTGCGCTTATGTACAATTAAGTACTACAGATCCGGGAATGAATTT 2115
Db |||||||
2605 GGCCTGTCTCAACTGCGCTTATGTACAATTAAGTACTACAGATCCGGGAATGAATTT 2546
QY 2116 GGACTACAG 2124
Db |||||||
2545 GGACTACAG 2537

RESULT 11
AF548377
LOCUS 297 bp DNA linear PLN 22-OCT-2002
DEFINITION Talaromyces emersonii oligosaccharyl transferase gene, partial cds.
ACCESSION AF548377
VERSION AF548377.1 GI:24210987
KEYWORDS
SOURCE Talaromyces emersonii
ORGANISM Talaromyces emersonii
REFERENCE
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; Talaromyces.
Crawley, K., Murray, P.G., Mc Loughlin, L.P. and Tuohy, M.G.
Cloning of oligosaccharyl transferase from *Talaromyces emersonii*
Unpublished
2 (bases 1 to 297)
Crawley, K., Murray, P.G., Mc Loughlin, L.P. and Tuohy, M.G.
Direct Submission
Submitted (23-SEP-2002) Biochemistry, National University of
Ireland Galway, Galway, Ireland
LOCATION/Qualifiers
1..297
/organism="Talaromyces emersonii"
/mol_type="genomic DNA"
/db_xref="taxon:68825"
1..297
/product="oligosaccharyl transferase"
1..297
/codon_start=1
/product="oligosaccharyl transferase"
/protein_id="AA52088.1"
/db_xref="GI:24210988"
/translation="GVIAPWGFRFSLMDTGAKIHPIIASVSEHPTAWPAFFDL
NFLVWLFPAQVYVCFDLKDEHFVIVIAVLAASYFAGVWVRLMLTIPVVCVAAA"

ORIGIN

Query Match 1.2%; Score 29; DB 8; Length 297;

Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1085 TAGCCCAAGATCCACATTCCTCCATTCATTC 1113
Db 52 TAGCCCAAGATCCACATTCCTCCATTCATTC 80

RESULT 12

AC118338
LOCUS 218635 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-167E4, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC118338
ACCESSION AC118338.3 GI:25138003
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 218635)
REFERENCE
AUTHORS
Muzny, D.M., Metzker, M.L., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Evans, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Hasland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louisedge, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martin, R., McNeill, T.Z.,
McNeilly, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankwelu, O., Okwuonu, G., Olariunpago, A., Pal, S., Parks, K.,
Pacternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L.,
Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Shen, H.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shasman, S., Shen, H.,
Shetty, J., Shivatsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, J., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmali, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von
Niederstock, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished

TITLE
JOURNAL

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (26-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Sep 26, 2002 this sequence version replaced gi:21736544.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GNXK
Center clone name: CH230-99P13
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 197284 bases at least Q40
Consensus quality: 203353 bases at least Q30
Consensus quality: 207410 bases at least Q20
Estimated insert size: 227312; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 69640: contig of 69640 bp in length
* 69641 69740: gap of unknown length
* 69741 81842: contig of 12102 bp in length
* 81843 81942: gap of unknown length
* 81943 86675: contig of 4733 bp in length
* 86676 86775: gap of unknown length
* 86776 97506: contig of 10731 bp in length
* 97507 97606: gap of unknown length
* 97607 136072: contig of 38466 bp in length
* 136073 136172: gap of unknown length
* 136173 257131: contig of 120959 bp in length
* 257132 257231: gap of unknown length
* 257232 259804: contig of 2573 bp in length
* 259805 259905: gap of unknown length
* 259906 263377: contig of 3473 bp in length.

FEATURES
source
1..263377
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-99P13"
86776..88251
/note="wgs_contig"
97607..99118
/note="wgs_contig"
143266..144894
/note="wgs_contig"
180092..181612
/note="wgs_contig"
220230..221437
/note="wgs_contig"
245895..249101
/note="wgs_contig"

misc_feature
86776..88251
/note="wgs_contig"
misc_feature
97607..99118
/note="wgs_contig"
misc_feature
143266..144894
/note="wgs_contig"
misc_feature
180092..181612
/note="wgs_contig"
misc_feature
220230..221437
/note="wgs_contig"
misc_feature
245895..249101
/note="wgs_contig"

ORIGIN

Query Match 1.2%; Score 29; DB 2; Length 263377;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2389 GAATAATACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 197394 GAATAATACCCCAAAAAAAAAAAAAAAAAAAAAA 197422

RESULT 14
AC116217
LOCUS
DEFINITION
AC116217
Rattus norvegicus clone CH230-111116, *** SEQUENCING IN PROGRESS
***, 7 unordered pieces.
AC116217
AC116217.3 GI:23664575
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 285045)
Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, K., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A.,
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, L.,
Devilla, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Detamo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensu, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Minet, G., Minje, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankwelu, O., Okwuonu, G., Olarinpo, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
Snead, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tirgey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
Weinhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 285045)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 3 (bases 1 to 285045)
 AUTHORS Rat Genome Sequencing Consortium.
 TITLE Direct Submission
 JOURNAL Submitted (10-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Oct 10, 2002 this sequence version replaced gi:21671715.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center Project name: GNLF
 Center Clone name: CH230-11110
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 240039 bases at least Q40
 Consensus quality: 244142 bases at least Q30
 Consensus quality: 246617 bases at least Q20
 Estimated insert size: 248451; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 129760: contig of 129760 bp in length
 129761 129860: gap of unknown length
 129861 129860: contig of 13848 bp in length
 268509 268608: gap of unknown length
 268609 278238: contig of 9630 bp in length
 278239 278338: gap of unknown length
 278339 279469: contig of 1131 bp in length
 279470 279569: gap of unknown length
 279570 281290: contig of 1721 bp in length
 281291 281390: gap of unknown length
 281391 282813: contig of 1423 bp in length
 282814 282913: gap of unknown length
 282914 285045: contig of 2132 bp in length.
 Location/Qualifiers
 1..285045
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clones="CH230-11110"
 1..1031
 /note="wgs_end_extension"

misc_feature
 clone_end:Sp6"
 1082..2099
 /note="wgs_end_extension"
 clone_end:Sp6"
 4359..5234
 /note="clone_boundary"
 clone_end:Sp6"
 site:EcoRI
 end_sequence:BH339077"
 4442..119103
 /note="clone_boundary"
 clone_end:Sp6"
 site:EcoRI
 end_sequence:BH339077"
 59773..76032
 /note="clone_boundary"
 clone_end:T7
 site:EcoRI
 end_sequence:BH339051"
 268609..270963
 /note="wgs_end_extension"
 clone_end:Sp6"
 ORIGIN
 Query Match 1.2%; Score 29; DB 2; Length 285045;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2389 GAATTAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
 Db 207074 GAATTAACCCCAAAAAAAAAAAAAAAAAAAAAA 207102

RESULT 15
 AK074363
 LOCUS AK074363 1968 bp mRNA PRI 12-SEP-2003
 DEFINITION Homo sapiens cDNA FLJ23783 fis, clone HEP21211, highly similar to
 Homo sapiens tripartite motif protein TRIM5 isoform alpha mRNA.
 ACCSSION AK074363
 VERSION AK074363.1 GI:18676945
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,
 Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
 Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
 NEDO human cDNA sequencing project
 TITLE Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 1968)
 AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
 Shibahara, T., Tanaka, T. and Nakamura, Y.
 Direct Submission
 TITLE Submitted (14-FEB-2002) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure, Human Genome
 Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail: flcdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
 Fax: 81-3-5449-5416)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan, cDNA full insert sequencing;
 Research Association for Biotechnology; cDNA library construction;
 5'- & 3'-end one pass sequencing; Department of Virology and Human
 Genome Center, Institute of Medical Science, University of Tokyo
 (partly supported by Science and Technology Agency).
 Location/Qualifiers
 1..1968
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HEP21211"
 /cell_line="HepG2"

FEATURES
 source
 1..1968
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HEP21211"
 /cell_line="HepG2"

```

/cell_type="hepatoma"
/clone_lib="HEP"
/note="Cloning vector: pME18SPL3"

```

ORIGIN

```

Query Match      1.2%; Score 28; DB 9; Length 1968;
Best Local Similarity 100.0%; Pred.No. 0.00077;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2390 AAATAAACCCAAAAA 2417
Db 1934 AAATAAACCCAAAAA 1961

```

Search completed: December 16, 2004, 00:55:29
Job time : 16238.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:53:33 ; Search time 1624 Seconds
(without alignments)
7812.721 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctagcgaagtgtgtcg.....ccaaaaaaaaaaaaaaaaa 2417

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004fs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 2417 | 100.0 | 2417 | 10 | ADD94789 | ADD94789 Drosophila |
| 2 | 2409 | 99.7 | 2699 | 4 | ABL13247 | ABL13247 Drosophila |
| 3 | 1463 | 60.5 | 4922 | 4 | ABL13246 | ABL13246 Drosophila |
| C 4 | 696 | 28.8 | 2953 | 4 | ABL18224 | ABL18224 Drosophila |
| C 5 | 249 | 10.3 | 2785 | 4 | ABL18208 | ABL18208 Drosophila |
| C 6 | 28 | 1.2 | 442 | 6 | ABQ59909 | ABQ59909 Human col |
| C 7 | 27 | 1.1 | 391 | 4 | AAL26051 | AAL26051 Human bre |
| C 8 | 27 | 1.1 | 401 | 4 | AAL08306 | AAL08306 Human bre |
| C 9 | 27 | 1.1 | 471 | 4 | AAL16849 | AAL16849 Human bre |
| C 10 | 26 | 1.1 | 2855 | 4 | ABL02795 | ABL02795 Drosophila |
| 11 | 26 | 1.1 | 3248 | 10 | ADC37655 | ADC37655 Human ang |
| 12 | 26 | 1.1 | 6153 | 4 | ABL02794 | ABL02794 Drosophila |
| C 13 | 25 | 1.0 | 253 | 5 | ABV61073 | ABV61073 Human pro |
| C 14 | 25 | 1.0 | 282 | 5 | ABV61368 | ABV61368 Human pro |
| C 15 | 25 | 1.0 | 340 | 5 | ABV43386 | ABV43386 Human pro |
| C 16 | 25 | 1.0 | 340 | 5 | ABV34528 | ABV34528 Human pro |
| C 17 | 25 | 1.0 | 354 | 6 | ABZ08446 | ABZ08446 Human leu |
| C 18 | 25 | 1.0 | 429 | 5 | ABV13410 | ABV13410 Human pro |
| C 19 | 25 | 1.0 | 434 | 12 | ADN13434 | ADN13434 Human pro |
| 20 | 25 | 1.0 | 513 | 5 | ADI75635 | ADI75635 Human ova |
| 21 | 25 | 1.0 | 513 | 5 | ADI69289 | ADI69289 Human ova |

| | | | | | | |
|------|----|-----|------|----|----------|---------------------|
| 22 | 25 | 1.0 | 575 | 4 | RAP33252 | Raf33252 Human sec |
| 23 | 25 | 1.0 | 583 | 5 | ADL40871 | Adl40871 Human ova |
| 24 | 25 | 1.0 | 605 | 5 | ABV58657 | Abv58657 Human pro |
| 25 | 25 | 1.0 | 659 | 6 | ABQ28115 | Abq28115 Oligonucle |
| C 26 | 25 | 1.0 | 659 | 6 | ABQ28114 | Abq28114 Oligonucle |
| 27 | 25 | 1.0 | 867 | 10 | ADG32855 | Adg32855 Human DNA |
| 28 | 25 | 1.0 | 1023 | 6 | ABT09673 | Abt09673 Human PAL |
| 29 | 25 | 1.0 | 1033 | 6 | ABT09672 | Abt09672 Human PAL |
| C 30 | 25 | 1.0 | 1033 | 6 | ABT09672 | Abt09672 Human PAL |
| C 31 | 25 | 1.0 | 8044 | 4 | AAS46625 | Aas46625 Tumour su |
| C 32 | 25 | 1.0 | 8244 | 10 | ADB54282 | Adb54282 Fretreat |
| C 33 | 24 | 1.0 | 130 | 5 | ABV61447 | Abv61447 Human pro |
| C 34 | 24 | 1.0 | 142 | 4 | PAL26330 | Pal26330 Human bre |
| C 35 | 24 | 1.0 | 143 | 5 | ADL44859 | Adl44859 Human ova |
| C 36 | 24 | 1.0 | 155 | 4 | RAL19280 | Ral19280 Human bre |
| C 37 | 24 | 1.0 | 167 | 5 | ADI76056 | Adi76056 Human ova |
| C 38 | 24 | 1.0 | 167 | 5 | ADI69719 | Adi69719 Human ova |
| C 39 | 24 | 1.0 | 175 | 8 | ABX53070 | Abx53070 Bovine ES |
| C 40 | 24 | 1.0 | 187 | 4 | AAL08585 | Aal08585 Human bre |
| C 41 | 24 | 1.0 | 200 | 6 | ABQ56626 | Abq56626 Human col |
| C 42 | 24 | 1.0 | 211 | 5 | ABV19351 | Abv19351 Human pro |
| C 43 | 24 | 1.0 | 216 | 5 | ABV61105 | Abv61105 Human pro |
| C 44 | 24 | 1.0 | 222 | 5 | ABV61623 | Abv61623 Human pro |
| 45 | 24 | 1.0 | 224 | 5 | RAS29077 | Ras29077 cDNA enco |

ALIGNMENTS

RESULT 1

ADD94789

ID ADD94789 standard; DNA; 2417 BP.

XX

AC ADD94789;

XX

DT 29-JAN-2004 (first entry)

XX

DE Drosophila melanogaster STT3 gene sequence.

XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;

KW major histocompatibility complex; human leukocyte antigen; HLA;

KW cytostatic; immunosuppressive; anticancer therapy; gene therapy; cancer;

KW lung cancer; intestine cancer; sarcoma; prostate cancer;

KW testicular cancer; breast cancer; melanoma; pancreatic cancer;

KW haematological cancer; immune response; lymphoid cell proliferation;

KW autoimmune disease; transplant rejection; SIMP-derived peptide;

XX fruit fly; gene; ds; STT3.

XX

OS Drosophila melanogaster.

XX

FN WC02003054008-A2.

XX

PD 03-JUL-2003.

XX

PF 18-DEC-2002; 2002WO-CA001967.

XX

PR 20-DEC-2001; 2001US-00028394.

XX

(COMP-) COMPATIGENE INC.

PA

PI Perreault C, McBride K;

XX

WPI; 2003-559122/52.

DR P-PSDB; ADD94790.

XX

PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic

PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung

PT or breast cancer, or for suppressing an immune response in an autoimmune

XX disease.

XX

PS Claim 6; SEQ ID NO 7; 66pp; English.

XX

CC This invention relates to a novel isolated or purified human protein.

CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the *Drosophila melanogaster* *Str3* gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Query Match 100.0%; Score 2417; DB 10; Length 2417;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAGAGTGTCTGCTTCATTCAGATCGCTTATATTTTCGAGTACTGGCTG 60
 DB 1 TCTAAGCGAGAGTGTCTGCTTCATTCAGATCGCTTATATTTTCGAGTACTGGCTG 60

QY 61 GAATTGGGACATGAATCGGACGGCGAAGATGCTGAACAGCAGAGTGGCTGGCTACAGCAG 120
 DB 61 GAATTGGGACATGAATCGGACGGCGAAGATGCTGAACAGCAGAGTGGCTGGCTACAGCAG 120

QY 121 CCTAATCATCTTCCGATCTGCTTATCGCTCGCTGGCGGAGTTTCCTCTCGCTCTT 180
 DB 121 CCTAATCATCTTCCGATCTGCTTATCGCTCGCTGGCGGAGTTTCCTCTCGCTCTT 180

QY 181 CGCGCTCATCGTTTCGAGTGGATTAATCCATGAGTTTGTATCCATGCTTCAACTACCGGC 240
 DB 181 CGCGCTCATCGTTTCGAGTGGATTAATCCATGAGTTTGTATCCATGCTTCAACTACCGGC 240

QY 241 CACCGCTCATCTGCTGAGATGTTGGTACACTTCTCACTGCTTCGAGCGGCGC 300
 DB 241 CACCGCTCATCTGCTGAGATGTTGGTACACTTCTCACTGCTTCGAGCGGCGC 300

QY 301 ATGATATCCGCTCGGCGAGATGTTGGCGGTACCGTCTATCCGCGCTGATGATTACGTC 360
 DB 301 ATGATATCCGCTCGGCGAGATGTTGGCGGTACCGTCTATCCGCGCTGATGATTACGTC 360

QY 361 CGGCGGATCCATTGGCTGCTGACGCTPACTCAACATCCGTCCTATTCGAGATCTG 420
 DB 361 CGGCGGATCCATTGGCTGCTGACGCTPACTCAACATCCGTCCTATTCGAGATCTG 420

QY 421 CGTGTCTGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
 DB 421 CGTGTCTGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

QY 481 GGAGCTGTGTCGCGGCGCGGCTTTCGCGCGGAGCTCATCGCATCGTCCGCTGG 540
 DB 481 GGAGCTGTGTCGCGGCGCGGCTTTCGCGCGGAGCTCATCGCATCGTCCGCTGG 540

QY 541 CTAATCAGTAGTGGTGGCTGGATCGTACGATGAACAGGAGCATTCGCTATTCGCT 600
 DB 541 CTAATCAGTAGTGGTGGCTGGATCGTACGATGAACAGGAGCATTCGCTATTCGCT 600

QY 601 GCAGTTTCACTTCTCTGCTGGTGGCTCAGTGAAGACTGGATCCGCTTCTGCTGGC 660
 DB 601 GCAGTTTCACTTCTCTGCTGGTGGCTCAGTGAAGACTGGATCCGCTTCTGCTGGC 660

QY 661 CGCAGCGCTTGTCTTCTTACATGTTGTCGCGCTGGGTTGCTTCTGCTTCAAT 720
 DB 661 CGCAGCGCTTGTCTTCTTACATGTTGTCGCGCTGGGTTGCTTCTGCTTCAAT 720

QY 721 CAACCTGATACCCCTGACGCTTCTGCTGCTTCTTATGCTGCTTATGCTGCTGCT 780

DB 721 CAACCTGATACCCCTGACGCTTCTGCTGCTTCTTATGCTGCTTATGCTGCTGCT 780

QY 781 GCTGACGAGTACAGCACTTCTTACATCCTGGGACTGCTGTTCTCCATGAGATCCGCT 840
 DB 781 GCTGACGAGTACAGCACTTCTTACATCCTGGGACTGCTGTTCTCCATGAGATCCGCT 840

QY 841 CGTGGGATTCACACGATACGACCACTGCAACATGCTGCTGGTGGGAGTGTGCT 900
 DB 841 CGTGGGATTCACACGATACGACCACTGCAACATGCTGCTGGTGGGAGTGTGCT 900

QY 901 CTTATGGCGGTGGCACTTCTGGCCATTTGCACTGCTGCTGCTGGCGCAACAGATTCCG 960
 DB 901 CTTATGGCGGTGGCACTTCTGGCCATTTGCACTGCTGCTGCTGGCGCAACAGATTCCG 960

QY 961 GAAGCTGTTTCATCTGCTGGCGGATTCCTGCTGGGCTTGGCTCTTTTGGCGCTGCTGGT 1020
 DB 961 GAAGCTGTTTCATCTGCTGGCGGATTCCTGCTGGGCTTGGCTCTTTTGGCGCTGCTGGT 1020

QY 1021 GCTCACATGCTGGGCTTGGCCCTGGGAGTGGAGTCTTCTTCTGCTGCTGGATAC 1080
 DB 1021 GCTCACATGCTGGGCTTGGCCCTGGGAGTGGAGTGGAGTCTTCTTCTGCTGCTGGATAC 1080

QY 1081 TGGCTACGCAAGATCCACATTCCTCATTTCCATCCGCTGCGAGCATCAGCCACAC 1140
 DB 1081 TGGCTACGCAAGATCCACATTCCTCATTTCCATCCGCTGCGAGCATCAGCCACAC 1140

QY 1141 TTGGTCTCTGCTTCTTCTTGTATCTGCACTCTGCTGCTGCTTCCAGTGGAGTGTG 1200
 DB 1141 TTGGTCTCTGCTTCTTCTTGTATCTGCACTCTGCTGCTGCTTCCAGTGGAGTGTG 1200

QY 1201 GCTACTCATCAAGAGATCAACAGCAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTG 1260
 DB 1201 GCTACTCATCAAGAGATCAACAGCAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTG 1260

QY 1261 GGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 DB 1261 GGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320

QY 1321 GCTGGCGGAGTGGCTTTCGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 DB 1321 GCTGGCGGAGTGGCTTTCGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380

QY 1381 GCGAATGGGACAGCAATAAGCGAGCAGCCAGCTGAGTGGATGAGCTGAGGATTCATTGA 1440
 DB 1381 GCGAATGGGACAGCAATAAGCGAGCAGCCAGCTGAGTGGATGAGCTGAGGATTCATTGA 1440

QY 1441 GAAGAGAGCTGTACGACAGGCTGGCAAGCTGAGCATCGTACTAAGCATGATGCCCA 1500
 DB 1441 GAAGAGAGCTGTACGACAGGCTGGCAAGCTGAGCATCGTACTAAGCATGATGCCCA 1500

QY 1501 GCAGGATATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
 DB 1501 GCAGGATATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

QY 1561 GTTGTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
 DB 1561 GTTGTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620

QY 1621 CATTTGCTTGGCTTTCACACAGTCAAGATGGATTCGCGCAACATTTAGACGATTCAG 1680
 DB 1621 CATTTGCTTGGCTTTCACACAGTCAAGATGGATTCGCGCAACATTTAGACGATTCAG 1680

QY 1681 AGAGGCTTACTTCTGCTTTCGAGAACACTGCGCATGCTGCTGCTGCTGCTGCTGCTG 1740
 DB 1681 AGAGGCTTACTTCTGCTTTCGAGAACACTGCGCATGCTGCTGCTGCTGCTGCTGCTG 1740

QY 1741 GGAATACGATACCGAGTAGCGGGAATGCGAAAACAGAACGAGCTAGTGGATTAATATAC 1800
 DB 1741 GGAATACGATACCGAGTAGCGGGAATGCGAAAACAGAACGAGCTAGTGGATTAATATAC 1800

QY 1801 GTGGAACAATAGTCTACATAGCGCTGTTGGGAGGCAATGCTTCAACCGAGGAGAGTC 1860

Db 1801 GTGGAACAATAGTACATAGCGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAGTC 1860
QY 1861 CTACGAAATATGACATCTTTGACGTGACTACGTTTGGTGATCTTTGGCGGTGAT 1920
Db 1861 CTACGAAATATGACATCTTTGACGTGACTACGTTTGGTGATCTTTGGCGGTGAT 1920
QY 1921 CGGCTATTCTGCGGATGATATCAACAAGTTCCTGTGGATGTCGGAATGCTGAGGGAGA 1980
Db 1921 CGGCTATTCTGCGGATGATATCAACAAGTTCCTGTGGATGTCGGAATGCTGAGGGAGA 1980
QY 1981 GCATCCCAAGGACATTAAGGAAGCGATTAATTTACCGACCGCGGTGAATTCAGGGTAGA 2040
Db 1981 GCATCCCAAGGACATTAAGGAAGCGATTAATTTACCGACCGCGGTGAATTCAGGGTAGA 2040
QY 2041 TGCCGAAGGTGCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATT 2100
Db 2041 TGCCGAAGGTGCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATT 2100
QY 2101 CGGGAATTAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAAAGCGGT 2160
Db 2101 CGGGAATTAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACAGTAAAGCGGT 2160
QY 2161 CATTTGGGAATTAAGGACTTCGATCTGACCTCTGAGGAGGCTTACACCAAGCAACTG 2220
Db 2161 CATTTGGGAATTAAGGACTTCGATCTGACCTCTGAGGAGGCTTACACCAAGCAACTG 2220
QY 2221 GCTTTGTCGATCTATAGGTTGAAGAGCCGATGATCAATAGACCATCACTGAAGAC 2280
Db 2221 GCTTTGTCGATCTATAGGTTGAAGAGCCGATGATCAATAGACCATCACTGAAGAC 2280
QY 2281 CAAGGAGAGAACGATTCCTCCAGCAAACTTCATTTCCAGAAAGAACTCTTAAGCGTCCAA 2340
Db 2281 CAAGGAGAGAACGATTCCTCCAGCAAACTTCATTTCCAGAAAGAACTCTTAAGCGTCCAA 2340
QY 2341 GGGCTACATAGCAAAACCGGTTGTTTAAAGGAAACGAACTTGAATAAACCCCA 2400
Db 2341 GGGCTACATAGCAAAACCGGTTGTTTAAAGGAAACGAACTTGAATAAACCCCA 2400
QY 2401 AAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAA 2417

RESULT 2

ABL13247
ID ABL13247 standard; cDNA; 2699 BP.
AC ABL13247;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US009231.

23-MAR-2000; 2000US-0191637P.

11-JUL-2000; 2000US-00614150.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

P-PSDB; ABB69144.

DR

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions.

PS Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Query Match 99.7%; Score 2409; DB 4; Length 2699;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2409; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTGTGCTTCATTTCCAGATCGGTTATAATTTTCGAGTTACTGGCTG 60

Db 10 TCTAAGCGAAGATGTGTGCTTCATTTCCAGATCGGTTATAATTTTCGAGTTACTGGCTG 69

QY 61 GAATTTGGGACATGATCGGACCGCCGAAGATGTCGAACGAAGTGGTGGCTACAGCAG 120

Db 70 GAATTTGGGACATGATCGGACCGCCGAAGATGTCGAACGAAGTGGTGGCTACAGCAG 129

QY 121 CTTAATCACTTCGGCATCTCTAATCGCTGGCTGGCGGATTTCTCTCGGCTCTT 180

Db 130 CTTAATCACTTCGGCATCTCTAATCGCTGGCTGGCGGATTTCTCTCGGCTCTT 189

QY 181 CCGCTCATCTGTTTCGAGTGCATTTCAATGAGTTTCAATCGTGGTTCAACTACCGGC 240

Db 190 CCGCTCATCTGTTTCGAGTGCATTTCAATGAGTTTCAATCGTGGTTCAACTACCGGC 249

QY 241 CACCGCTTACATGTTGTCAGAAATGTTGTCACACTTCTCACTGGTTCGAGGAGCGGC 300

Db 250 CACCGCTTACATGTTGTCAGAAATGTTGTCACACTTCTCACTGGTTCGAGGAGCGGC 309

QY 301 ATGGTATCCGCTCGGAGGATTTGGCGGTACCGTCTATCCCGGCTGTGATTCGTC 360

Db 310 ATGGTATCCGCTCGGAGGATTTGGCGGTACCGTCTATCCCGGCTGTGATTCGTC 369

QY 361 CGGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG 420

Db 370 CGGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG 429

QY 421 CGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480

Db 430 CGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 489

QY 481 GGAGCTGTGTTCCGCGGCGCGGCTCTTCGCGCGAGTTCATCCGCTGCGTGG 540

Db 490 GGAGCTGTGTTCCGCGGCGCGGCTCTTCGCGCGAGTTCATCCGCTGCGTGG 549

QY 541 CTACATGATAGTTCGGTGGCTGATCGTAGAAGAGGCAATTCGCATATTCGCGCT 600

Db 550 CTACATGATAGTTCGGTGGCTGATCGTAGAAGAGGCAATTCGCATATTCGCGCT 609

QY 601 GCAGTTCACTTCTGTTGGTGGCTCAGTGAAGACTGGATCCGTTCTTGGTGGC 660

Db 610 GCAGTTCACTTCTGTTGGTGGCTCAGTGAAGACTGGATCCGTTCTTGGTGGC 669

QY 661 GCGAGCGCTTGTCTTACTTCTACATGTTGTCGCGCTGGGCTGCTACGTGTTTCAT 720

Db 670 GCGAGCGCTTGTCTTACTTCTACATGTTGTCGCGCTGGGCTGCTACGTGTTTCAT 729

QY 721 CAACCTGATACCCCTGCAACGCTTTCGTACTGCTCATTTATGGGAGGAGTACTCGCGCGTCT 780

Db 730 CAACCTGATACCCCTGACGCTCTTCGTAAGTCTCAATATGGCGAGGTACTCGCGCGTCT 789
Qy 781 GCTGACCACTACAGCACTTCTACATCCCTGGAGTCTGTTCTCCATGCAAGATCCCTTT 840
Db 790 GCTGACCACTACAGCACTTCTACATCCCTGGAGTCTGTTCTCCATGCAAGATCCCTTT 849
Qy 841 CCGTGGATTCACACGATACGACCAAGTGAACACATGGCTGCGCTGGAGTCTTGTGCT 900
Db 850 CCGTGGATTCACACGATACGACCAAGTGAACACATGGCTGCGCTGGAGTCTTGTGCT 909
Qy 901 CCTTATGCGCTGGCGCACTTGGCGCAATTTGCACTGCGCTGCTCGCGCAAGAGTTCG 960
Db 910 CCTTATGCGCTGGCGCACTTGGCGCAATTTGCACTGCGCTGCTCGCGCAAGAGTTCG 969
Qy 961 GAAGCTGTTCACTGCTGGCGCAATTTGCTGGTGGCGTGGCGTCTTGTGGCGCTGCTGT 1020
Db 970 GAAGCTGTTCACTGCTGGCGCAATTTGCTGGTGGCGTGGCGTCTTGTGGCGCTGCTGT 1029
Qy 1021 GCTCACCATGCTGGCGCTGGCGCAATTTGCTGGTGGCGTGGCGTCTTGTGGCGCTGCTGT 1080
Db 1030 GCTCACCATGCTGGCGCTGGCGCAATTTGCTGGTGGCGTGGCGTCTTGTGGCGCTGCTGT 1089
Qy 1081 TGGCTACGCCAAGATCCACATTCCTCATTCATTCGATTCGCTGGAGCATCAGCCACCAC 1140
Db 1090 TGGCTACGCCAAGATCCACATTCCTCATTCATTCGATTCGCTGGAGCATCAGCCACCAC 1149
Qy 1141 TTGGTCTCGTCTTCTTCTTCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 1200
Db 1150 TTGGTCTCGTCTTCTTCTTCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 1209
Qy 1201 GPACTGCATGACGAGATCAACGACGAGCGGTTTCGTTGCTGCTGACGCCATCAGTGC 1260
Db 1210 GPACTGCATGACGAGATCAACGACGAGCGGTTTCGTTGCTGCTGACGCCATCAGTGC 1269
Qy 1261 GGTTCATCTCGTGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1270 GGTTCATCTCGTGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
Qy 1321 GCTGCGCGAGTGGCTTTTCGGAATGTTGGATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1380
Db 1330 GCTGCGCGAGTGGCTTTTCGGAATGTTGGATGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1389
Qy 1381 GCGATGCGCAGACCAATGCGCAGCAGCCAGCGAGTGGATGAGCTGAGGATTCATTGA 1440
Db 1390 GCGATGCGCAGACCAATGCGCAGCAGCCAGCGAGTGGATGAGCTGAGGATTCATTGA 1449
Qy 1441 GAAGAAGCGTGTACGACAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCCA 1500
Db 1450 GAAGAAGCGTGTACGACAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCCA 1509
Qy 1501 GCAGATGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1510 GCAGATGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1569
Qy 1561 GTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1570 GTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629
Qy 1621 CATTCGTTGCTTCCACACAGTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1630 CATTCGTTGCTTCCACACAGTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 1689
Qy 1681 AGAGCTTACTGCTGCTTCCAGACATCTGCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1690 AGAGCTTACTGCTGCTTCCAGACATCTGCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1749
Qy 1741 GGATTCGGATACAGATGAGCGGAATGGCAACAGAGCAGCTAGTGGATTAATATAC 1800
Db 1750 GGATTCGGATACAGATGAGCGGAATGGCAACAGAGCAGCTAGTGGATTAATATAC 1809
Qy 1801 GTGGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860

Db 1810 GTGGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1869
Qy 1861 CTACGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 1870 CTACGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
Qy 1921 CGGCTATTTCTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
Db 1930 CGGCTATTTCTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1989
Qy 1981 GCATCCCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
Db 1990 GCATCCCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2049
Qy 2041 TGCCGAAGTGTCTCGGCGCTGCTCAAGTCTTATGATGATGATGATGATGATGATGATGATGAT 2100
Db 2050 TGCCGAAGTGTCTCGGCGCTGCTCAAGTCTTATGATGATGATGATGATGATGATGATGATGAT 2109
Qy 2101 CGGGAATTTGAAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
Db 2110 CGGGAATTTGAAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2169
Qy 2161 CATGGGAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
Db 2170 CATGGGAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2229
Qy 2221 GCTTGTTCGATCTATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
Db 2230 GCTTGTTCGATCTATAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2289
Qy 2281 CAAGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Db 2290 CAAGGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2349
Qy 2341 GGCTATCATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
Db 2350 GGCTATCATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2409
Qy 2401 AAAAAAAAAA 2409
Db 2410 AAAAAAAAAA 2418

RESULT 3
ABL3246/c
ID ABL3246 standard; cDNA; 4922 BP.
XX
XX ABL3246;
XX AC
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX CS Drosophila melanogaster.
XX FN WO200171042-A2.
XX PD
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR P-FSD5; ABB69143.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 34220; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB146175) and the encoded proteins (ABBS7737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
Query Match 60.5%; Score 1463; DB 4; Length 4922;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAACGAGATGTCGTGCAATTCAGATCGTTATATATTTTCGAGTTACTGCTG 60
DB 3913 TCTAACGAGATGTCGTGCAATTCAGATCGTTATATATTTTCGAGTTACTGCTG 3854
QY 61 GAATTCGGACATGAATCGGACCCGGAAGATGCTGAACAGCAAGGTGGCTACAGCAG 120
DB 3853 GAATTCGGACATGAATCGGACCCGGAAGATGCTGAACAGCAAGGTGGCTACAGCAG 3794
QY 121 CTTAATCAGCTTCGCGATCCTGTAATCGCTGCGTGGCGGATTTCTCTCGCCTTT 180
DB 3793 CTTAATCAGCTTCGCGATCCTGTAATCGCTGCGTGGCGGATTTCTCTCGCCTTT 3734
QY 181 CGCGCTATCCGTTTCGAGTCGATTATCCATGAGTTTCATCCGCTGTTCAACTACCGGCG 240
DB 3733 CGCGCTATCCGTTTCGAGTCGATTATCCATGAGTTTCATCCGCTGTTCAACTACCGGCG 3674
QY 241 CACCGCTACATGTTGCAAGTGGTGTGACAACTTCCTCACTGTTCCGACGAGCGGC 300
DB 3673 CACCGCTACATGTTGCAAGTGGTGTGACAACTTCCTCACTGTTCCGACGAGCGGC 3614
QY 301 ATGGTATCCGCTCGGAGGATGTCGGCGGTACCGTCTATCCCGCCTGATGATTACGTC 360
DB 3613 ATGGTATCCGCTCGGAGGATGTCGGCGGTACCGTCTATCCCGCCTGATGATTACGTC 3554
QY 361 CGCGGAATCCATTCGCTGCTGACGATCAATACGATCGGTCCTATTCGTGACATCTG 420
DB 3553 CGCGGAATCCATTCGCTGCTGACGATCAATACGATCGGTCCTATTCGTGACATCTG 3494
QY 421 CGTGTTCTGGGCGGATCTTCAGTGGCTGACCTCCATCCACCTACCTGCTGACCAA 480
DB 3493 CGTGTTCTGGGCGGATCTTCAGTGGCTGACCTCCATCCACCTACCTGCTGACCAA 3434
QY 481 GAGCTGTGGTCCGCGGCGCGGCTCTTCGCGCGAGCTTCATCCGCTACGCTGCG 540
DB 3433 GAGCTGTGGTCCGCGGCGCGGCTCTTCGCGCGAGCTTCATCCGCTACGCTGCG 3374
QY 541 CTACATCAGTAGTTCGGTGGCTGATCGTACGATTAACGAGGCGATTGTCATATTCGCGCT 600
DB 3373 CTACATCAGTAGTTCGGTGGCTGATCGTACGATTAACGAGGCGATTGTCATATTCGCGCT 3314
QY 601 GCAGTTCACTACTTCCTGTTGGTGGCTCAGTGAAGTGGATCCGTTCTGTCGGC 660
DB 3313 GCAGTTCACTACTTCCTGTTGGTGGCTCAGTGAAGTGGATCCGTTCTGTCGGC 3254
QY 661 CGCAGCGCTTTGCTCTACTTCTCATGTTGTCGCGCTGGGCTGCTACGTTTCATCAT 720
DB 3253 CGCAGCGCTTTGCTCTACTTCTCATGTTGTCGCGCTGGGCTGCTACGTTTCATCAT 3194
QY 721 CAACCTGATACCCCTGCACTCTTCGTACTGCTCATTTATGGCAGGTAAGTCTCGCGCTCT 780

Db 3193 CAACCTGATACCCCTGCACTCTTCGTACTGCTCATTTATGGGAGGTAAGTCTCGCGCTCT 3134
QY 781 GGTACCAAGCTACAGCACTCTTCTACATCCTGGGACTGCTGTTCTTCCATGAGATCCCTT 840
Db 3133 GGTACCAAGCTACAGCACTCTTCTACATCCTGGGACTGCTGTTCTTCCATGAGATCCCTT 3074
QY 841 CGTGGATTCCTCAACCGATACGACAGTGAACACATCGCTGCGTGGGAGTGTGCT 900
Db 3073 CGTGGATTCCTCAACCGATACGACAGTGAACACATCGCTGCGTGGGAGTGTGCT 3014
QY 901 CTTATGCGCTGGGCGACCTTTGCGCCATTTGCACTCCGCTCTGTCGCGCAACAGTTCG 960
Db 3013 CTTATGCGCTGGGCGACCTTTGCGCCATTTGCACTCCGCTCTGTCGCGCAACAGTTCG 2954
QY 961 GAAGCTGTTCACTGTCGCGCGATTTGCTGGGCTGGGCTCTTTGTCGCGCTGCTGCT 1020
Db 2953 GAAGCTGTTCACTGTCGCGCGATTTGCTGGGCTGGGCTCTTTGTCGCGCTGCTGCT 2894
QY 1021 GCTACCATGTCGGCGCTTTGCGCCCTGAGTGGACGCTTTCTATCTGCTGGGATAC 1080
Db 2893 GCTACCATGTCGGCGCTTTGCGCCCTGAGTGGACGCTTTCTATCTGCTGGGATAC 2834
QY 1081 TGGCTACGCGCAAGATCCACATTTCCCATCTGTCGATCGTTCGAGCATCAGCCACAC 1140
Db 2833 TGGCTACGCGCAAGATCCACATTTCCCATCTGTCGATCGTTCGAGCATCAGCCACAC 2774
QY 1141 TTGGTTCGCTTCTTTTGTGATCTGCACTCTGCTGTCGCTTCCCATGCGGAGTGTG 1200
Db 2773 TTGGTTCGCTTCTTTTGTGATCTGCACTCTGCTGTCGCTTCCCATGCGGAGTGTG 2714
QY 1201 GTACTGCACTACGAGATCAACGAGCGCGTTCGTCGTCGTCGTCGTCGTCGTCGTCG 1260
Db 2713 GTACTGCACTACGAGATCAACGAGCGCGTTCGTCGTCGTCGTCGTCGTCGTCGTCG 2654
QY 1261 GGTTCATCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 2653 GGTTCATCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2594
QY 1321 GCTGCGCGAGTGGGCTTTTCGGGACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 2593 GCTGCGCGAGTGGGCTTTTCGGGACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2534
QY 1381 GCGAATGGGCAAGCCTAAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
Db 2533 GCGAATGGGCAAGCCTAAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2474
QY 1441 GAAGAAGAGCTGTACGACAGG 1463
Db 2473 GAAGAAGAGCTGTACGACAGG 2451

RESULT 4
ABL18224/c
ID ABL18224 standard; DNA; 2953 BP.
XX
AC ABL18224;
XX
DT 26-MAR-2002 (first entry)
XX
DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 6145.
XX
KW *Drosophila*; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX *Drosophila melanogaster*.
OS
XX WO2001/1042-A2.
XX
XX 27-SBP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX

PR 23-MAR-2000; 2000US-0191637P.
 XX 11-JUL-2000; 2000US-00614150.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 6145; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2953 BP; 924 A; 628 C; 648 G; 753 T; 0 U; 0 Other;
 Query Match 28.8%; Score 696; DB 4; Length 2953;
 Best Local Similarity 100.0%; Pred. No. 1e-255;
 Matches 696; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCTAAGCGAAGATGTGCTGCTGATTCAGATCGGTATATTTTCGAGTTACTGGCTG 60
 DB 696 TCTAAGCGAAGATGTGCTGCTGATTCAGATCGGTATATTTTCGAGTTACTGGCTG 637
 QY 61 GAATTTGGACATGAATCGGACCGGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
 DB 636 GAATTTGGACATGAATCGGACCGGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 577
 QY 121 CCTAATCACTTCGCAATCCGCTGAATCGCTGGCTGGCGGATTTCTCTCGGCTCTT 180
 DB 576 CTAATCACTTCGCAATCCGCTGAATCGCTGGCTGGCGGATTTCTCTCGGCTCTT 517
 QY 181 CGCGTCACTCGTTTCGAGTCGATTCATCCATGATTTGATCCGCTGTTCAACTACGGGC 240
 DB 516 CGCGTCACTCGTTTCGAGTCGATTCATCCATGATTTGATCCGCTGTTCAACTACGGGC 457
 QY 241 CACCGCTACATGTGCGAGATGTTGGTACAACTTCCTCACTGTTGCGAGGGGC 300
 DB 456 CACCGCTACATGTGCGAGATGTTGGTACAACTTCCTCACTGTTGCGAGGGGC 397
 QY 301 ATGGTATCCGTCGCGAGGATTTGGCGGATACCGTCTATCCCGGCTGATGATACGTC 360
 DB 396 ATGGTATCCGTCGCGAGGATTTGGCGGATACCGTCTATCCCGGCTGATGATACGTC 337
 QY 361 CGCGGATTCATTTGGCTGCTGACGCTACTCAACATACCGTCCATATTCGATCTG 420
 DB 336 CGCGGATTCATTTGGCTGCTGACGCTACTCAACATACCGTCCATATTCGATCTG 277
 QY 421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
 DB 276 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 217
 QY 481 CGAGCTGTGTCGCGGCGCGCTCTTTCGCGCGAGCTTCATGCCATGTCGCTGG 540
 DB 216 CGAGCTGTGTCGCGGCGCGCTCTTTCGCGCGAGCTTCATGCCATGTCGCTGG 157
 QY 541 CTACATCAGTAGGTGGCTGGATCTGATACGATACAGGGGATTCGCCATATTCGCCCT 600
 DB 156 CTACATCAGTAGGTGGCTGGATCTGATACGATACAGGGGATTCGCCATATTCGCCCT 97
 QY 601 GCAGTTCACTTCTCTGTGGGTGCGTCACTGAGACTGGATCGGTCTTCTGGTCGCG 660

DB 96 GCAGTTCACTTCTCTGTGGGTGCGTCACTGAGACTGATCCGTTCTGTGTCGCG 37
 QY 661 CGCAGCGGCTTTGCTCTACTTCTACATGTTGTCGCG 696
 DB 36 CGCAGCGGCTTTGCTCTACTTCTACATGTTGTCGCG 1
 RESULT 5
 ABL18208/c.
 ID ABL18208 standard; DNA; 2785 BP.
 XX
 AC ABL18208;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6097.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 FS Claim 1; SEQ ID NO 6097; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2785 BP; 795 A; 644 C; 604 G; 742 T; 0 U; 0 Other;
 Query Match 10.3%; Score 249; DB 4; Length 2785;
 Best Local Similarity 100.0%; Pred. No. 3.2e-85;
 Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1876 ATCTCTTACGTGAGTACGTTTGTGATCTTTTGGCGGTGTGATCGCTATTCGCGA 1935
 DB 2785 ATCTCTTACGTGAGTACGTTTGTGATCTTTTGGCGGTGTGATCGCTATTCGCGA 2726
 QY 1936 TGATATCAACAGTTCCTGTCGATGTCGATTCCTGAGGAGAGATCCCAAGGACAT 1995
 DB 2725 TGATATCAACAGTTCCTGTCGATGTCGATTCCTGAGGAGAGATCCCAAGGACAT 2666
 QY 1996 TAAGGAAAGCGATTACTTTTACCGACCGCGGTGAATTCAGGGTAGATCCGAAGTGTCC 2055
 DB 2665 TAAGGAAAGCGATTACTTTTACCGACCGCGGTGAATTCAGGGTAGATCCGAAGTGTCC 2606

QY 2056 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTCGGGAATGAAGTT 2115
Db 2605 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTCGGGAATGAAGTT 2546
QY 2116 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTCGGGAATGAAGTT 2546
Db 2545 GGCCTGCTCACTCCCTATGTACAAATTAAGCTACTACAGATTCGGGAATGAAGTT 2546

RESULT 6
ABQ59909/c
ID ABQ59909 standard; cDNA; 442 BP.
XX
AC ABQ59909;
XX
DT 02-AUG-2002 (first entry)
XX
DE Human colon cancer related nucleotide sequence SEQ ID NO:3604.
XX
KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX genetic analysis; diagnostic; antisense therapy; gene; ss.
XX Homo sapiens.
XX
FN WO200229086-A2.
XX
PD 11-APR-2002.
XX

PF 02-OCT-2001; 2001WO-US030732.
XX
XX 02-OCT-2000; 2000US-0237271P.
XX
PA (FARB) BAYER CORP.
XX
PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI Thiagalingam A, Lewis ME;
XX
DR WPI: 2002-426115/45.
XX
XX New isolated nucleic acid that is differentially expressed in cancer
PT tissues useful for determining the presence of colon cancer in a cell or
PT tissue type, and in antisense therapy.
XX

PS Claim 1; Fig 1; 796pp; English.
XX
CC ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins
CC encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC used in antisense therapy. An antibody immunoreactive with a polypeptide
CC encoded by (I) is useful for detecting cancer in a patient sample, and
CC for detecting the presence or absence of a polynucleotide encoded by a
CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridizes to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists
XX
SQ Sequence 442 BP; 102 A; 102 C; 91 G; 132 T; 0 U; 15 Other;

Query Match 1.2%; Score 28; DB 6; Length 442;
Best Local Similarity 100.0%; Pred.No. 0.87;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AAATAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 34 AAATAAACCCCAAAAAAAAAAAAAAAAAAAAAA 7

RESULT 7
AAL26051/c
ID AAL26051 standard; cDNA; 391 BP.
XX
AC AAL26051;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18508.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
FN WO200151628-A2.
XX
PD 19-JUL-2001.
XX

PF 10-JAN-2001; 2001WO-US000798.
XX
XX 14-JAN-2000; 2000US-0176077P.
XX
PR 14-MAR-2000; 2000US-0189167P.
XX
PR 24-MAR-2000; 2000US-0192099P.
XX
PR 29-MAR-2000; 2000US-0193480P.
XX
PR 09-JUN-2000; 2000US-0205230P.
XX
PR 25-JUL-2000; 2000US-0211315P.
XX
PR 25-JUL-2000; 2000US-0220534P.
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI: 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer.

XX Claim 1; Page 3414; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity

XX Sequence 391 BP; 96 A; 65 C; 70 G; 160 T; 0 U; 0 Other;

Query Match 1.1%; Score 27; DB 4; Length 391;

Best Local Similarity 100.0%; Pred.No. 2.1;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AAATAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2416

Db 29 AAATAAACCCCAAAAAAAAAAAAAAAAAAAAAA 3

RESULT 8

AAL08306/c

ID AAL08306 standard; cDNA; 401 BP.

XX

AC AAL08306;

XX 07-DEC-2001 (first entry)

XX Human breast cancer expressed polynucleotide 763.

XX Human; breast cancer; cell marker; cytostatic; ss.

XX Homo sapiens.

XX

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737,
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;
 Query Match 1.1%; Score 26; DB 4; Length 2855;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1937 GATATCAACAAGTTCCTGTGGATGGT 1962
 Db 2213 GATATCAACAAGTTCCTGTGGATGGT 2238
 |||||

RESULT 11
 ADC37655
 ID ADC37655 standard; cDNA; 3248 BP.
 AC ADC37655;
 XX
 DT 18-DEC-2003 (first entry)
 DE Human angiotensin-like protein 1b encoding cDNA SEQ ID NO:4.
 XX human; angiotensin-like protein 1; AMLP1; cytostatic; gene therapy;
 KW AMLP1b; gene; ss.
 OS Homo sapiens.
 PN WO2003037931-A2.
 XX
 PD 08-MAY-2003.
 PF 01-NOV-2002; 2002WO-US035129.
 XX
 PR 01-NOV-2001; 2001US-0334773P.
 PA (AMSH) AVERSHAM BIOSCIENCES SV CORP.
 XX
 PI Shannon M, Phan T;
 DR WPI; 2003-430501/40.
 DR P-PSDB; ADC37657.
 XX
 PT New isolated nucleic acid molecule encoding a human angiotensin-like
 PT protein, useful for treating or preventing a disorder associated with
 PT decreased or increased expression or activity of AMLP1.
 PS Claim 1; SEQ ID NO 4; 172pp; English.

XX The present invention describes the human angiotensin-like protein 1
 CC (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene
 CC therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and
 CC compositions of the present invention can be used for treating or
 CC preventing a disorder associated with decreased or increased expression
 CC or activity of AMLP1. the present sequence encodes human AMLP1b from the
 CC present invention.

XX SQ Sequence 3248 BP; 955 A; 865 C; 880 G; 548 T; 0 U; 0 Other;
 Query Match 1.1%; Score 26; DB 10; Length 3248;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 ATAAACCCCAAAAAAAAAAAAAAAAAA 2417
 Db 3207 ATAAACCCCAAAAAAAAAAAAAAAAAA 3232
 |||||

RESULT 12

ABL02794
 ID ABL02794 standard; cDNA; 6153 BP.
 XX
 AC ABL02794;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2864.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-658660/75.
 DR P-PSDB; ABB58691.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PS Claim 1; SEQ ID NO 2864; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737,
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 6153 BP; 1692 A; 1334 C; 1299 G; 1828 T; 0 U; 0 Other;
 Query Match 1.1%; Score 26; DB 4; Length 6153;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1937 GATATCAACAAGTTCCTGTGGATGGT 1962
 Db 4511 GATATCAACAAGTTCCTGTGGATGGT 4536
 |||||

RESULT 13
 ABB61073/c
 ID ABB61073 standard; cDNA; 253 BP.
 XX
 AC ABB61073;
 XX
 DT 13-SEP-2002 (first entry)
 DE Human prostate expression marker cDNA 61064.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.

PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 11603; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 253 BP; 41 A; 69 C; 46 G; 97 T; 0 U; 0 Other;
 Query Match 1.0%; Score 25; DB 5; Length 253;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
 Db 29 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 5
 RESULT 14
 ABV61368/c
 ID ABV61368 standard; cDNA; 282 BP.
 XX
 AC ABV61368;
 XX
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 61359.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 11603; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 253 BP; 41 A; 69 C; 46 G; 97 T; 0 U; 0 Other;
 Query Match 1.0%; Score 25; DB 5; Length 253;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
 Db 29 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 5

PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 PS Claim 1; Page 11641; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 282 BP; 69 A; 66 C; 61 G; 86 T; 0 U; 0 Other;
 Query Match 1.0%; Score 25; DB 5; Length 282;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
 Db 42 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 18
 RESULT 15
 ABV43386/c
 ID ABV43386 standard; cDNA; 340 BP.
 XX
 AC ABV43386;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 43377.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX

XX

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 8649; 11750pp; English.

XX

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the specification or its complement. (I) is useful as (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) also useful as a pharmacodynamic or pharmacogenomic marker ; (I) is also useful as a

XX

Sequence 340 BP; 97 A; 46 C; 63 G; 134 T; 0 U; 0 Other;

Query Match

Query Match 1.0%; Score 25; DB 5; Length 340;

Best Local Similarity 100.0%; Pred. No. 13;

Qv

2393 TAAACCCAAATAAAAAAAAAA 2417

୩୩

84 TAAACCCAAATAAAATAAAATAAA 60

Search completed: December 15, 2004, 15:47:00
Job time : 1627 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:34:04 ; Search time 253.148 Seconds
(without alignments)
6786.440 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctaaagcgaagaatgtgtcg.....ccaaaaaAAAAAAAAAAAA 2417

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 889 | 36.8 | 900 | 4 | US-09-270-767-12331 |
| 2 | 755 | 31.2 | 1660 | 4 | US-09-270-767-158 |
| 3 | 755 | 31.2 | 1660 | 4 | US-09-270-767-15440 |
| C 4 | 503 | 20.8 | 560 | 4 | US-09-270-767-1318 |
| C 5 | 503 | 20.8 | 560 | 4 | US-09-270-767-16600 |
| C 6 | 299 | 12.4 | 1660 | 4 | US-09-270-767-158 |
| C 7 | 299 | 12.4 | 1660 | 4 | US-09-270-767-15440 |
| 8 | 133 | 5.5 | 133 | 4 | US-09-270-767-38082 |
| 9 | 26 | 1.1 | 914 | 4 | US-09-270-767-12856 |
| 10 | 24 | 1.0 | 632 | 3 | US-09-385-982-499 |
| 11 | 24 | 1.0 | 660 | 1 | US-08-555-678-41 |
| 12 | 24 | 1.0 | 1060 | 4 | US-08-997-685A-7 |
| 13 | 24 | 1.0 | 1182 | 3 | US-08-261-637-261 |
| 14 | 24 | 1.0 | 1391 | 1 | US-08-261-652-1 |
| 15 | 24 | 1.0 | 1391 | 5 | PCR-US95-0752-1 |
| 16 | 24 | 1.0 | 1720 | 4 | US-09-823-038A-41 |
| 17 | 24 | 1.0 | 4137 | 3 | US-09-498-964-2 |
| 18 | 24 | 1.0 | 4302 | 3 | US-09-245-281-38 |
| 19 | 24 | 1.0 | 4302 | 4 | US-09-207-359B-38 |
| 20 | 24 | 1.0 | 4302 | 4 | US-09-340-620A-38 |
| 21 | 24 | 1.0 | 4302 | 4 | US-09-865-364-38 |
| 22 | 23 | 1.0 | 132 | 4 | US-09-621-976-16776 |
| 23 | 23 | 1.0 | 134 | 4 | US-09-621-976-16780 |
| 24 | 23 | 1.0 | 135 | 4 | US-09-621-976-16778 |
| 25 | 23 | 1.0 | 139 | 4 | US-09-621-976-16785 |
| 26 | 23 | 1.0 | 140 | 4 | US-09-621-976-16783 |
| 27 | 23 | 1.0 | 140 | 4 | US-09-621-976-16786 |

| | | | | | | |
|------|----|-----|------|---|---------------------|-------------------|
| 28 | 23 | 1.0 | 147 | 4 | US-09-270-767-3687 | Sequence 3687, Ap |
| 29 | 23 | 1.0 | 147 | 4 | US-09-270-767-18969 | Sequence 18969, A |
| 30 | 23 | 1.0 | 157 | 4 | US-09-621-976-16781 | Sequence 16781, A |
| 31 | 23 | 1.0 | 190 | 4 | US-09-621-976-16784 | Sequence 16784, A |
| 32 | 23 | 1.0 | 383 | 4 | US-09-621-976-12067 | Sequence 12067, A |
| 33 | 23 | 1.0 | 383 | 4 | US-09-621-976-12237 | Sequence 12237, A |
| 34 | 23 | 1.0 | 383 | 4 | US-09-621-976-13563 | Sequence 13563, A |
| 35 | 23 | 1.0 | 453 | 4 | US-09-621-976-2916 | Sequence 2916, Ap |
| 36 | 23 | 1.0 | 1071 | 4 | US-09-205-258-118 | Sequence 118, App |
| 37 | 23 | 1.0 | 1133 | 4 | US-09-205-258-236 | Sequence 236, App |
| 38 | 23 | 1.0 | 1316 | 4 | US-09-370-838-30 | Sequence 30, Appl |
| 39 | 23 | 1.0 | 1316 | 4 | US-08-854-133-30 | Sequence 30, Appl |
| 40 | 23 | 1.0 | 2030 | 3 | US-08-706-216-3 | Sequence 3, Appl |
| 41 | 23 | 1.0 | 2030 | 4 | US-09-650-284B-3 | Sequence 3, Appl |
| 42 | 23 | 1.0 | 2254 | 4 | US-09-138-277C-2 | Sequence 2, Appl |
| 43 | 23 | 1.0 | 3467 | 4 | US-09-743-492A-1 | Sequence 1, Appl |
| C 44 | 23 | 1.0 | 5455 | 4 | US-10-204-708-33 | Sequence 33, Appl |
| 45 | 23 | 1.0 | 7702 | 4 | US-09-023-655-1336 | Sequence 1336, Ap |

ALIGNMENTS

RESULT 1

US-09-270-767-12331

; Sequence 12331, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 12331

; LENGTH: 900

; TYPE: DNA

; ORGANISM: Drosophila melanogaster

US-09-270-767-12331

Query Match 36.8%; Score 889; DB 4; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 889; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | TCTAAGCAGAAATGTGTGTTGCATTTTCAGATCGGTATATAATTTTCGATTACTGGCTG | 60 |
| DB | 12 | TCTAAGCAGAAATGTGTGTTGCATTTTCAGATCGGTATATAATTTTCGATTACTGGCTG | 71 |
| QY | 61 | GAATTTGGGACATGAATCGGACGCGAAGATGCTGAACAGCAAGTGTGCTGGCTACAGCAG | 120 |
| DB | 72 | GAATTTGGGACATGAATCGGACGCGAAGATGCTGAACAGCAAGTGTGCTGGCTACAGCAG | 131 |
| QY | 121 | CCTAATACCTTCCCATCTCTGCTTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT | 180 |
| DB | 132 | CCTAATACCTTCCCATCTCTGCTTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT | 191 |
| QY | 181 | CGCGTCATCGTTTCGAGTCGATTATCCATGATTTGATCCGTTGTTCACTACCGGCG | 240 |
| DB | 192 | CGCGTCATCGTTTCGAGTCGATTATCCATGATTTGATCCGTTGTTCACTACCGGCG | 251 |
| QY | 241 | CACCGCTACATGTCGAGAAATGGTTGGTACAATCTTCTCAATGGTTGCAAGAGCGCG | 300 |
| DB | 252 | CACCGCTACATGTCGAGAAATGGTTGGTACAATCTTCTCAATGGTTGCAAGAGCGCG | 311 |
| QY | 301 | ATGTTATCCGCTCGCGAGGATTGTCGCGGTACCGTCTATCCCGGCTGATGATTACGTC | 360 |
| DB | 312 | ATGTTATCCGCTCGCGAGGATTGTCGCGGTACCGTCTATCCCGGCTGATGATTACGTC | 371 |
| QY | 361 | CGCGGAATCAATGGCTGCTGCAACGTAATCAACATACCGGTCCATATTCGTGACATCG | 420 |
| DB | 372 | CGCGGAATCAATGGCTGCTGCAACGTAATCAACATACCGGTCCATATTCGTGACATCG | 431 |

| | | | |
|----|-----|---|-----|
| Qy | 421 | CGTGTCTCTGGCGCGGATCTTCACTGGCCCTGACCTCCATCTCCACATCACTGCTGACCAA | 480 |
| Db | 432 | CGTGTCTCTGGCGCGGATCTTCACTGGCCCTGACCTCCATCTCCACCTACTGCTGACCAA | 491 |
| Qy | 481 | GGAGCTGTGGTCCGGGGGCGCGGCTCTTCGGCGGCGACCTTCATCGCCATCGTGCCTGG | 540 |
| Db | 492 | GGAGCTGTGGTCCGGGGGCGCGGCTCTTCGGCGGCGACCTTCATCGCCATCGTGCCTGG | 551 |
| Qy | 541 | CTACATCAGTAGTGGTGGTGGATCGTAGCAATACGAGGGCAATGGCCATATTCGCCCT | 600 |
| Db | 552 | CTACATCAGTAGTGGTGGTGGATCGTAGCAATACGAGGGCAATGGCCATATTCGCCCT | 611 |
| Qy | 601 | GCAGTTCACCTACTTCTGTGGGTGGCGCTCAGTGAAGATGGATCCGTGTTCGTGGTCGC | 660 |
| Db | 612 | GCAGTTCACCTACTTCTGTGGGTGGCGCTCAGTGAAGATGGATCCGTGTTCGTGGTCGC | 671 |
| Qy | 661 | CGCAGCGCTTTGTCTACTTCTACATGGTGTCCGCTGGGTGGGTACGTGTTCATCAT | 720 |
| Db | 672 | CGCAGCGCTTTGTCTACTTCTACATGGTGTCCGCTGGGTGGGTACGTGTTCATCAT | 731 |
| Qy | 721 | CAACCTGATACCCCTGCACGCTTCGTACTGCTCATTTAGGGCAGGTACTCCGCGCTCT | 780 |
| Db | 732 | CAACCTGATACCCCTGCACGCTTCGTACTGCTCATTTAGGGCAGGTACTCCGCGCTCT | 791 |
| Qy | 781 | GCTGACCACTACAGCACTTCTACATCTCTGGGACCTGCTGTCTCCATGCAGATCCCTTT | 840 |
| Db | 792 | GCTGACCACTACAGCACTTCTACATCTCTGGGACCTGCTGTCTCCATGCAGATCCCTTT | 851 |
| Qy | 841 | CGTGGGATTCAAACCGATACGCAACAGTGAACACATGGCTGGCTGGGA | 899 |
| Db | 852 | CGTGGGATTCAAACCGATACGCAACAGTGAACACATGGCTGGCTGGGA | 900 |

RESULT 2

```

US-09-270-767-158
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1995-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

```

| | | | |
|------|----|--|------|
| 1028 | Db | GTTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGCGGAGATATCAACAAGATTCTCTG | 1087 |
| 1955 | QY | TGGATGTCGGAATTCTGAGGAGAGCATCCCAAGGACATTAAAGGAAACGGATTACTTT | 2014 |
| 1098 | Db | TGGATGTCGGAATTCTGAGGAGAGCATCCCAAGGACATTAAAGGAAACGGATTACTTT | 1147 |
| 2015 | QY | ACCGACCGCGTGAATTCAGGGTGTAGATGCCGAAGGTGCTCCGGCCCTGCTCACTGCCCT | 2074 |
| 1148 | Db | ACCGACCGCGTGAATTCAGGGTGTAGATGCCGAAGGTGCTCCGGCCCTGCTCACTGCCCT | 1207 |
| 2075 | QY | ATGTACAAATTAAGCTACTACAGATTTCGGGGAAATTTGAAGTTTGGACTACAGAGTCCATCT | 2134 |
| 1208 | Db | ATGTACAAATTAAGCTACTACAGATTTCGGGGAAATTTGAAGTTTGGACTACAGAGTCCATCT | 1267 |
| 2135 | QY | GGATATCATCGCACACGTTAAGCCGCTCATTTGGGAATAAGGACTTTCGATCTTGACCTACCTG | 2194 |
| 1268 | Db | GGATATCATCGCACACGTTAAGCCGCTCATTTGGGAATAAGGACTTTCGATCTTGACCTACCTG | 1327 |
| 2195 | QY | GAGGAGCCTACACACAGAACACTGGGCTTGTTCGCATCTATAGGGTGAAAGACCGCAT | 2254 |
| 1328 | Db | GAGGAGCCTACACACAGAACACTGGGCTTGTTCGCATCTATAGGGTGAAAGACCGCAT | 1387 |
| 2255 | QY | GAGTTCAATAGACCATCACTGAAGACCAAGGAGAGACGATTTCCTCAGCAAACTTCAIT | 2314 |
| 1388 | Db | GAGTTCAATAGACCATCACTGAAGACCAAGGAGAGACGATTTCCTCAGCAAACTTCAIT | 1447 |
| 2315 | QY | TCGGAAGAAGACTTAAGGCTGCGAAGGCTACATACGAACCCGACCGGTTGTGTTAAG | 2374 |
| 1448 | Db | TCGGAAGAAGACTTAAGGCTGCGAAGGCTACATACGAACCCGACCGGTTGTGTTAAG | 1507 |
| 2375 | QY | GGAAACCGAACCTTGAATAAACCACCAAAAAA | 2409 |
| 1508 | Db | GGAAACCGAACCTTGAATAAACCACCAAAAAA | 1542 |

RESULT 3

US-09-270-767-15440

/ Sequence 15440, Application US/09270767

/ Patent No. 6703491

/ GENERAL INFORMATION:

/ APPLICANT: Homburger et al.

/ TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

/ FILE REFERENCE: File Reference: 7326-094

/ CURRENT APPLICATION NUMBER: US/09/270,767

/ CURRENT FILING DATE: 1999-03-17

/ NUMBER OF SEQ ID NOS: 62517

/ SOFTWARE: Patentin Ver. 2.0

/ SEQ ID NO 15440

/ LENGTH: 1660

/ TYPE: DNA

/ ORGANISM: *Drosophila melanogaster*

US-09-270-767-15440

RESULT 3

```

US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1998-03-17
; NUMBER OF SEQ ID NOS: 63517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

```

QY 1895 GTTTTGGTGAATCTTTGGCGGTGTGATCGGCTATTCTCGCGATGATATCAACAAGTTCTTG 1954
Db 1028 GTTTTGGTGAATCTTTGGCGGTGTGATCGGCTATTCTCGCGATGATATCAACAAGTTCTTG 1087
QY 1955 TGGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTT 2014
Db 1088 TGGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTT 1147
QY 2015 ACCGACCGGGTGAATTCAGGGTAGATCCGAAGTGTCTCGGCGCTCTCACTGCTT 2074
Db 1148 ACCGACCGGGTGAATTCAGGGTAGATCCGAAGTGTCTCGGCGCTCTCACTGCTT 1207
QY 2075 ATGTACAAATTAAGCTACTACAGATTTCGGGAAATTGAAGTTGGACTACAGAGGTCCATCT 2134
Db 1208 ATGTACAAATTAAGCTACTACAGATTTCGGGAAATTGAAGTTGGACTACAGAGGTCCATCT 1267
QY 2135 GATATGATCGACACGTAAAGCGGTCAATTCGGGATTAAGGACTTCGATCTGACTACCTG 2194
Db 1268 GATATGATCGACACGTAAAGCGGTCAATTCGGGATTAAGGACTTCGATCTGACTACCTG 1327
QY 2195 GAGGAGGCTACACACAGAACACTTGGCTTGTTCGCATCTATAGGGTGAAGAAAGCCGCAT 2254
Db 1328 GAGGAGGCTACACACAGAACACTTGGCTTGTTCGCATCTATAGGGTGAAGAAAGCCGCAT 1387
QY 2255 GATTCATATAGACCTACTAGAACCAAGAGAGACGATTCCTCAGCAAACTTCATT 2314
Db 1388 GATTCATATAGACCTACTAGAACCAAGAGAGACGATTCCTCAGCAAACTTCATT 1447
QY 2315 TCGAGAAAGAACTTAAGCGTCGCAAGGGCTACATAGCAAAACCGGTTGTTGTTAAG 2374
Db 1448 TCGAGAAAGAACTTAAGCGTCGCAAGGGCTACATAGCAAAACCGGTTGTTGTTAAG 1507
QY 2375 GGAACACGACCTTGAAATTAACCCAAAAA 2409
Db 1508 GGAACACGACCTTGAAATTAACCCAAAAA 1542

RESULT 4

US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1318
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Query Match 20.8%; Score 503; DB 4; Length 560;

Best Local Similarity 99.8%; Pred. No. 7.2e-208;

Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1107 TCATTGCATCCGTGTCGAGCATCAGCCACCACTTGGTCTCGTCTCTTTGATCTGC 1166
Db 560 TCATTGCATCCGTGTCGAGCATCAGCCACCACTTGGTCTCGTCTCTTTGATCTGC 501
QY 1167 ACATCTGTTGTCGCTTCCAGTGGAGTGTGGTACTGATCAAGCAGATCAAGCAGC 1226
Db 500 ACATCTGTTGTCGCTTCCAGTGGAGTGTGGTACTGATCAAGCAGATCAAGCAGC 441
QY 1227 AGCGGTTTTTCGTGTTGCTGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1286
Db 440 AGCGGTTTTTCGTGTTGCTGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 381
QY 1287 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1466
Db 1287 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1346

Db 380 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 321
QY 1347 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1406
Db 320 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 261
QY 1407 CCACCGAGTGGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1466
Db 260 CCACCGAGTGGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 201
QY 1467 GCAAGCTGAAGCATCTGATCAAGCATGATCCAGCAGGATATCTGCGTCAAGTCCAAAC 1526
Db 200 GCAAGCTGAAGCATCTGATCAAGCATGATCCAGCAGGATATCTGCGTCAAGTCCAAAC 141
QY 1527 TGAAGAGTATGTTTATTTTCGCGTCTTAATGTTGATGATGTTGCTGCTGCTCACTGCA 1586
Db 140 TGAAGAGTATGTTTATTTTCGCGTCTTAATGTTGATGATGTTGCTGCTGCTCACTGCA 81
QY 1587 CGTGGTGACAGCAATGCTTCTCAAGTCCCTCAATGTTGCTTGGCTTTCCCAAGCAGTC 1646
Db 80 CGTGGTGACAGCAATGCTTCTCAAGTCCCTCAATGTTGCTTGGCTTTCCCAAGCAGTC 21
QY 1647 AAGATGGATCCCGC 1660
Db 20 AAGATGGATCCCGC 7

RESULT 5

US-09-270-767-1660/c

; Sequence 1660, Application US/09270767
; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1660
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1660

Query Match 20.8%; Score 503; DB 4; Length 560;

Best Local Similarity 99.8%; Pred. No. 7.2e-208;

Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1107 TCATTGCATCCGTGTCGAGCATCAGCCACCACTTGGTCTCGTCTCTTTGATCTGC 1166
Db 560 TCATTGCATCCGTGTCGAGCATCAGCCACCACTTGGTCTCGTCTCTTTGATCTGC 501
QY 1167 ACATCTGTTGTCGCTTCCAGTGGAGTGTGGTACTGATCAAGCAGATCAAGCAGC 1226
Db 500 ACATCTGTTGTCGCTTCCAGTGGAGTGTGGTACTGATCAAGCAGATCAAGCAGC 441
QY 1227 AGCGGTTTTTCGTGTTGCTGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1286
Db 440 AGCGGTTTTTCGTGTTGCTGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 381
QY 1287 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1346
Db 380 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 321
QY 1347 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1406
Db 320 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 261
QY 1407 CCACCGAGTGGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1466
Db 1407 CCACCGAGTGGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 1346
QY 1466 GTTTCATGTTGATGAGTTCGATCGGCAATCAGTGGGTTTACTTCGCTGGTGTGATGTGC 201

QY 1467 GCAAGCTGAAGCATCTACTAAGCATGATGCCAGCAGGATGATGCGCTCAGCTCCACCC 1526
DB 200 GCAAGCTGAAGCATCTACTAAGCATGATGCCAGCAGGATGATGCGCTCAGCTCCACCC 141
QY 1527 TGAAGAGTATTGTTTATTTGCGCGTCTTAATGCTGTGTGATGATTTTCGCTGCTCACTGCA 1586
DB 140 TGAAGAGTATTGTTTATTTGCGCGTCTTAATGCTGTGTGATGATTTTCGCTGCTCACTGCA 81
QY 1587 CGTGGGTGACCAAGTCCCTACTCCAGTCCCTCCATGCTCTGGGCTTCCACACAGTCC 1646
DB 80 CGTGGGTGACCAAGTCCCTACTCCAGTCCCTCCATGCTCTGGGCTTCCACACAGTCC 21
QY 1647 AAGATGATCCCGC 1660
DB 20 AAGATGATCCCGC 7

RESULT 6

US-09-270-767-158/c
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

Query Match 12.4%; Score 299; DB 4; Length 1660;
Best Local Similarity 99.1%; Pred. No. 1e-119;
Matches 549; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1760 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAACTAGTCACATA 1819
DB 793 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAACTAGTCACATA 734
QY 1820 GCGCTGGTTGGCAAGCAATGCTTTCAACCGAGGAGAGTCTTACGAAATTAATGACATCT 1879
DB 733 GCGCTGGTTGGCAAGCAATGCTTTCAACCGAGGAGAGTCTTACGAAATTAATGACATCT 674
QY 1880 CTTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 1939
DB 673 CTTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 614
QY 1940 ATCAACAACTTCTGTGATGCTCGGAATTTAGCTTACGAGGAGATCCCAAGGACATTAAG 1999
DB 613 ATCAACAACTTCTGTGATGCTCGGAATTTAGCTTACGAGGAGATCCCAAGGACATTAAG 554
QY 2000 GAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATGCGGAGTGTCTCCGGCC 2059
DB 553 GAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATGCGGAGTGTCTCCGGCC 494
QY 2060 CTGCTCAACTGCTTATGTACAAATTAAGCTTACAGATTCGGGGAATTAAGTTGGAC 2119
DB 493 CTGCTCAACTGCTTATGTACAAATTAAGCTTACAGATTCGGGGAATTAAGTTGGAC 434
QY 2120 TACAGAGTCCATCTGGATATGATCGCACACGTAACCGCGTCAITGGGAATAAGGACTTC 2179
DB 433 TACAGAGTCCATCTGGATATGATCGCACACGTAACCGCGTCAITGGGAATAAGGACTTC 374
QY 2180 GATCTGACCTTACCTGGAGGAGGCTTACACACAGACACTGGCTTGTTCGATCTATAGG 2239
DB 373 GATCTGACCTTACCTGGAGGAGGCTTACACACAGACACTGGCTTGTTCGATCTATAGG 314
QY 2240 GTGAAGAGCCGCTAGTTCATTAAGACCATCTCTGAAGACCAAGGAGAGAAAGGATTCCT 2299

RESULT 8

US-09-270-767-28082
; Sequence 28082, Application US/09270767

DB 313 GTGAAGAGCCGCTAGTTCATATAGACCATCATCTGAGACCAAGGAGAGACGATTCCT 254
QY 2300 CCAGCAAACTTCAT 2313
DB 253 CCAGCAAACTTCAT 240
RESULT 7
US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 82517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Query Match 12.4%; Score 299; DB 4; Length 1660;
Best Local Similarity 99.1%; Pred. No. 1e-119;
Matches 549; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1760 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAACTAGTCACATA 1819
DB 793 GCGGGAATGGCAACAGACGACCTAGTGGATAATATACGTGGAACTAGTCACATA 734
QY 1820 GCGCTGGTTGGCAAGCAATGCTTTCAACCGAGGAGAGTCTTACGAAATTAATGACATCT 1879
DB 733 GCGCTGGTTGGCAAGCAATGCTTTCAACCGAGGAGAGTCTTACGAAATTAATGACATCT 674
QY 1880 CTTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 1939
DB 673 CTTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGAT 614
QY 1940 ATCAACAACTTCTGTGATGCTCGGAATTTAGCTTACGAGGAGATCCCAAGGACATTAAG 1999
DB 613 ATCAACAACTTCTGTGATGCTCGGAATTTAGCTTACGAGGAGATCCCAAGGACATTAAG 554
QY 2000 GAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATGCGGAGTGTCTCCGGCC 2059
DB 553 GAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATGCGGAGTGTCTCCGGCC 494
QY 2060 CTGCTCAACTGCTTATGTACAAATTAAGCTTACAGATTCGGGGAATTAAGTTGGAC 2119
DB 493 CTGCTCAACTGCTTATGTACAAATTAAGCTTACAGATTCGGGGAATTAAGTTGGAC 434
QY 2120 TACAGAGTCCATCTGGATATGATCGCACACGTAACCGCGTCAITGGGAATAAGGACTTC 2179
DB 433 TACAGAGTCCATCTGGATATGATCGCACACGTAACCGCGTCAITGGGAATAAGGACTTC 374
QY 2180 GATCTGACCTTACCTGGAGGAGGCTTACACACAGACACTGGCTTGTTCGATCTATAGG 2239
DB 373 GATCTGACCTTACCTGGAGGAGGCTTACACACAGACACTGGCTTGTTCGATCTATAGG 314
QY 2240 GTGAAGAGCCGCTAGTTCATTAAGACCATCTCTGAAGACCAAGGAGAGAAAGGATTCCT 2299
DB 313 GTGAAGAGCCGCTAGTTCATTAAGACCATCTCTGAAGACCAAGGAGAGAAAGGATTCCT 254
QY 2300 CCAGCAAACTTCAT 2313
DB 253 CCAGCAAACTTCAT 240

; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 28082
; LENGTH: 133
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-28082

Query Match 5.5%; Score 133; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 6.5e-48;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 757 TATGGCAGGTACTCGCGCGTCTGCTGACGAGCTACAGCACTTCTACATCTCTGGGACT 816
Db 1 TATGGCAGGTACTCGCGCGTCTGCTGACGAGCTACAGCACTTCTACATCTCTGGGACT 60

QY 817 GCTGTTCTCCATGCAGATCCCTTCTGTTGGATTCCACCGATACGACCACTGGAACACAT 876
Db 61 GCTGTTCTCCATGCAGATCCCTTCTGTTGGATTCCACCGATACGACCACTGGAACACAT 120

QY 877 GCTGCGCTGGGA 889
Db 121 GCTGCGCTGGGA 133

RESULT 9
US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 12856
; LENGTH: 914
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-12856

Query Match 1.1%; Score 26; DB 4; Length 914;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1937 GATATCAACAAGTCTCTGTGGATGTT 1962
Db 272 GATATCAACAAGTCTCTGTGGATGTT 297

RESULT 10
US-09-385-982-499
; Sequence 499, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 499
; LENGTH: 632
; TYPE: DNA
; ORGANISM: *Homo sapiens*
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(632)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-499

Query Match 1.0%; Score 24; DB 3; Length 632;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2394 AAACCCAAAAA 2417
Db 505 AAACCCAAAAA 528

RESULT 11
US-08-555-678-41
; Sequence 41, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,678
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,459
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49DUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-555-678-41

Query Match 1.0%; Score 24; DB 1; Length 660;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 114 AAACCCAAAAA 137

RESULT 12

US-08-997-685A-7/c
; Sequence 7, Application US/08997685A
; Patent No. 6551821
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University
; APPLICANT: Kandel, Eric
; TITLE OF INVENTION: Brain Cyclic Nucleotide Gated Ion Channel and Uses Thereof
; FILE REFERENCE: 0575/54806
; CURRENT APPLICATION NUMBER: US/08/997,685A
; CURRENT FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: mouse;
US-08-997-685A-7

Query Match 1.0%; Score 24; DB 4; Length 1060;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 28 AAACCCAAAAA 5

RESULT 13

US-09-461-697-261
; Sequence 261, Application US/09461697
; Patent No. 6277574
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Furman, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-261

Query Match 1.0%; Score 24; DB 3; Length 1182;
Best Local Similarity 100.0%; Pred. No. 0.76;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 1150 AAACCCAAAAA 1173

RESULT 14

US-08-261-662-1
; Sequence 1, Application US/08261662
; Patent No. 5523227
; GENERAL INFORMATION:
; APPLICANT: Bram, Richard J
; APPLICANT: Crabtree, Gerald R
; TITLE OF INVENTION: CELL CALCIUM REGULATION AND ITS USE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,662
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-59392/BIR STAN-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1391 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 37..927
US-08-261-662-1

Query Match 1.0%; Score 24; DB 1; Length 1391;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2394 AAACCCAAAAA 2417
Db 1343 AAACCCAAAAA 1366

RESULT 15

US-09-505-0752-1
; Sequence 1, Application PC/TUS9507752
; GENERAL INFORMATION:
; APPLICANT: The Board of Trustees of the Leland Stanford Junior
; APPLICANT: University
; TITLE OF INVENTION: CELL CALCIUM REGULATION AND ITS USE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/07752
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rowland, Bertram I
/ REGISTRATION NUMBER: 20,015
/ REFERENCE/DOCKET NUMBER: A-59392/BIR STAN-167
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1391 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 37..927
PCT-US95-07752-1
```

```
Query Match 1.0%; Score 24; DB 5; Length 1391;
Best Local Similarity 100.0%; Pred.No. 0.75;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 2394 AAACCCAAAAAAAAAAAAAAAAAAAA 2417
Db 1343 AAACCCAAAAAAAAAAAAAAAAAAAA 1366
```

```
Search completed: December 16, 2004, 07:21:31
Job time : 255.148 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2004, 00:55:34 ; Search time 2138.68 Seconds
(without alignments)
6239.285 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctagcgaagaatgtctg.....ccaaaaa.....2417

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 4093002 seqs, 2760418825 residues

Word size : 0

Total number of hits satisfying chosen parameters: 8186004

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 2417 | 100.0 | 2417 | 15 | US-10-028-384-7 |
| C 2 | 28 | 1.2 | 442 | 11 | Sequence 3604, Ap |
| C 3 | 26 | 1.1 | 354 | 18 | Sequence 67624, A |
| C 4 | 26 | 1.1 | 395 | 14 | Sequence 1610, Ap |
| C 5 | 26 | 1.1 | 348 | 18 | Sequence 4, Appl |
| C 6 | 25 | 1.0 | 153 | 16 | Sequence 53154, A |
| C 7 | 25 | 1.0 | 153 | 16 | Sequence 53154, A |
| C 8 | 25 | 1.0 | 293 | 16 | Sequence 14946, A |
| C 9 | 25 | 1.0 | 293 | 16 | Sequence 14946, A |
| C 10 | 25 | 1.0 | 354 | 16 | Sequence 8437, Ap |
| C 11 | 25 | 1.0 | 416 | 16 | Sequence 44394, A |
| C 12 | 25 | 1.0 | 416 | 16 | Sequence 44394, A |

| | | | | | | |
|------|----|-----|-------|----|----------------------|--------------------|
| C 13 | 25 | 1.0 | 488 | 18 | US-10-425-115-78691 | Sequence 78691, A |
| C 14 | 25 | 1.0 | 513 | 10 | US-09-814-353-2031 | Sequence 2031, Ap |
| C 15 | 25 | 1.0 | 513 | 10 | US-09-814-353-8377 | Sequence 8377, Ap |
| C 16 | 25 | 1.0 | 572 | 16 | US-10-191-803-934 | Sequence 934, Ap |
| C 17 | 25 | 1.0 | 583 | 10 | US-09-814-353-14761 | Sequence 14761, A |
| C 18 | 25 | 1.0 | 592 | 18 | US-10-425-115-114851 | Sequence 114851, A |
| C 19 | 25 | 1.0 | 600 | 16 | US-10-424-599-99562 | Sequence 99562, A |
| C 20 | 25 | 1.0 | 659 | 18 | US-10-363-345A-14705 | Sequence 14705, A |
| C 21 | 25 | 1.0 | 659 | 18 | US-10-363-345A-14706 | Sequence 14706, A |
| C 22 | 25 | 1.0 | 679 | 18 | US-10-425-115-47797 | Sequence 47797, A |
| C 23 | 25 | 1.0 | 1023 | 9 | US-09-804-682-24 | Sequence 24, Appl |
| C 24 | 25 | 1.0 | 1033 | 9 | US-09-804-682-23 | Sequence 23, Appl |
| C 25 | 25 | 1.0 | 2335 | 15 | US-10-341-434-237 | Sequence 237, Ap |
| C 26 | 25 | 1.0 | 5244 | 18 | US-10-473-126-312 | Sequence 312, Ap |
| C 27 | 25 | 1.0 | 8044 | 16 | US-10-321-714A-347 | Sequence 347, Ap |
| C 28 | 25 | 1.0 | 23611 | 17 | US-10-322-281-61 | Sequence 61, Appl |
| C 29 | 24 | 1.0 | 139 | 17 | US-10-021-323-5739 | Sequence 5739, Ap |
| C 30 | 24 | 1.0 | 143 | 10 | US-09-814-353-18749 | Sequence 18749, A |
| C 31 | 24 | 1.0 | 143 | 16 | US-10-242-535A-51365 | Sequence 51365, A |
| C 32 | 24 | 1.0 | 143 | 16 | US-10-085-783A-51365 | Sequence 51365, A |
| C 33 | 24 | 1.0 | 167 | 10 | US-09-814-353-2461 | Sequence 2461, Ap |
| C 34 | 24 | 1.0 | 167 | 10 | US-09-814-353-8798 | Sequence 8798, Ap |
| C 35 | 24 | 1.0 | 175 | 9 | US-09-983-965-2999 | Sequence 2999, Ap |
| C 36 | 24 | 1.0 | 180 | 16 | US-10-424-599-69408 | Sequence 69408, A |
| C 37 | 24 | 1.0 | 200 | 11 | US-09-969-034-321 | Sequence 321, Appl |
| C 38 | 24 | 1.0 | 224 | 9 | US-09-784-846-58 | Sequence 58, Appl |
| C 39 | 24 | 1.0 | 224 | 14 | US-10-091-483-58 | Sequence 58, Appl |
| C 40 | 24 | 1.0 | 229 | 16 | US-10-062-674-971 | Sequence 971, Ap |
| C 41 | 24 | 1.0 | 243 | 16 | US-10-242-535A-16977 | Sequence 16977, A |
| C 42 | 24 | 1.0 | 243 | 16 | US-10-085-783A-16977 | Sequence 16977, A |
| C 43 | 24 | 1.0 | 244 | 18 | US-10-425-115-10205 | Sequence 10205, A |
| C 44 | 24 | 1.0 | 247 | 18 | US-10-425-115-148698 | Sequence 148698, A |
| C 45 | 24 | 1.0 | 251 | 18 | US-10-425-115-166919 | Sequence 166919, A |

ALIGNMENTS

RESULT 1
US-10-028-384-7
Sequence 7, Application US/10028384
Publication No. US20030148285A1
GENERAL INFORMATION:
APPLICANT: COMPATIGENE INC.
APPLICANT: PERREAU, Claude
APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028,384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent in version 3.1
SEQ ID NO 7
LENGTH: 2417
TYPE: DNA
ORGANISM: Drosophila melanogaster
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: AF132552
DATABASE ENTRY DATE: 1999-04-27
RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

Query Match 100.0%; Score 2417; DB 15; Length 2417;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGCGAAGAAATGTGCTTGCATTTCAGATCGGTATAATTTTCGAGTTACTGGCTG 60
DB 1 TCTAGCGAAGAAATGTGCTTGCATTTCAGATCGGTATAATTTTCGAGTTACTGGCTG 60
QY 61 GAATTTGGACATGAATCGGACCGCCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
DB 61 GAATTTGGACATGAATCGGACCGCCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120

121 CCTAATCACTTCGCGCATCTCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT 180
 121 CCTAATCACTTCGCGCATCTCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTT 180
 181 CGCGTCACTCGCTTCGAGTCGATTCATCCATGAGTTTGTATCGTGGTTCATTCGCGGC 240
 181 CGCGTCACTCGCTTCGAGTCGATTCATCCATGAGTTTGTATCGTGGTTCATTCGCGGC 240
 241 CACCGCTTACATGCTGCGAGATGCTGGTACATCTCTCACTCTCACTCTCACTCTCACTCT 300
 241 CACCGCTTACATGCTGCGAGATGCTGGTACATCTCTCACTCTCACTCTCACTCTCACTCT 300
 301 ATGGTATCCGCTCGCGAGATGCTGGTACATCTCTCACTCTCACTCTCACTCTCACTCT 360
 301 ATGGTATCCGCTCGCGAGATGCTGGTACATCTCTCACTCTCACTCTCACTCTCACTCT 360
 361 CGCGGAAATCCATGCTGCTGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 361 CGCGGAAATCCATGCTGCTGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 421 CGTGTTCCTGCGCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 421 CGTGTTCCTGCGCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 481 GGAGTGTGCTGCGCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 481 GGAGTGTGCTGCGCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 541 CTACATCAGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 541 CTACATCAGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 601 CGAGTTCACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 601 CGAGTTCACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 661 CGAGTTCACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 661 CGAGTTCACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 721 CAACCTGATACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 721 CAACCTGATACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 781 GCTGACAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 781 GCTGACAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 841 CGTGGGATTCACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 841 CGTGGGATTCACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 901 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 901 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
 961 GAAGTGTTCATGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 961 GAAGTGTTCATGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 1021 GCTCACCATGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 1021 GCTCACCATGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 1081 TGGCTACCGGATTCACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 1081 TGGCTACCGGATTCACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 1141 TTGGTTCCTGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 1141 TTGGTTCCTGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

1201 GTACTGCTCAAGCAGATCAACGACGCGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
 1201 GTACTGCTCAAGCAGATCAACGACGCGGCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
 1261 GGTTTACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 1261 GGTTTACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 1321 GCTGCGCGAGTGGCTTTTTCGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 1321 GCTGCGCGAGTGGCTTTTTCGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 1381 GCGAATGGGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
 1381 GCGAATGGGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
 1441 GAAGAAGCCTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
 1441 GAAGAAGCCTGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1500
 1501 GCAGGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 1501 GCAGGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 1561 GTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 1561 GTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 1621 CATTTGCTTGGCTTTTCCACACAGTCAAGTGGATCCCGCAACATTTTGTAGCAGTTCAG 1680
 1621 CATTTGCTTGGCTTTTCCACACAGTCAAGTGGATCCCGCAACATTTTGTAGCAGTTCAG 1680
 1681 AGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 1681 AGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 1741 GGATTAACGATACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 1741 GGATTAACGATACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 1801 GTGGAACATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 1801 GTGGAACATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 1861 CTACGAAATTAATGATCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 1861 CTACGAAATTAATGATCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 1921 CGGCTATTTCTGCGGATGATACCAAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 1921 CGGCTATTTCTGCGGATGATACCAAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 1981 GCATCCCAAGGACATTAAGGAAAGGATTTACTTTACCGACCGGCTGAAATTCAGGGTAGA 2040
 1981 GCATCCCAAGGACATTAAGGAAAGGATTTACTTTACCGACCGGCTGAAATTCAGGGTAGA 2040
 2041 TGCAGAGGCTGCTCGGCGCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
 2041 TGCAGAGGCTGCTCGGCGCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
 2101 CGGGGAATTAAGTGGATACAGAGTTCATCTGGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 2101 CGGGGAATTAAGTGGATACAGAGTTCATCTGGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 2161 CATTTGGATTAAGGACTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 2220
 2161 CATTTGGATTAAGGACTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 2220
 2221 GCTTGTTCGCTATATAGGGTGAAGAGCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 2221 GCTTGTTCGCTATATAGGGTGAAGAGCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 2281 CRAAGGAGAGACGATTTCTCTCCAGCAAACTTCTTCTTCTGAGAAAGAACTCTTAAGCGTCCGAA 2340

Db 2281 CAAGGAGAGAACGATTCCTCCAGCAAACTTCATTTCCGAGAGAACTCTAGACGTCGCAA 2340
QY 2341 GGGCTACATACAGAACCGCGTGTGTTAAGGGAAGAACGACCTTGAAATAAACCCA 2400
Db 2341 GGGCTACATACAGAACCGCGTGTGTTAAGGGAAGAACGACCTTGAAATAAACCCA 2400
QY 2401 AAAAAAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAAAAAA 2417

RESULT 2

US-09-969-034-3604/c
; Sequence 3604, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poonima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; TITLE OF INVENTION: Expressed in Cancer Tissue
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3604
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 89, 220, 278, 287, 337, 350, 378, 391, 404, 406, 422, 432,
; LOCATION: 434, 435, 442
; OTHER INFORMATION: n = A,T,C or G
US-09-969-034-3604

Query Match 1.2%; Score 28; DB 11; Length 442;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2390 AAATAAACCCCAAAAAAAAAAAAAAAAAA 2417
Db 34 AAATAAACCCCAAAAAAAAAAAAAAAAAA 7

RESULT 3

US-10-425-115-67624/c
; Sequence 67624, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 67624
; LENGTH: 354
; TYPE: DNA
; ORGANISM: Zea mays

; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_16166C.1
US-10-425-115-67624

Query Match 1.1%; Score 26; DB 18; Length 354;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 ATAAACCCCAAAAAAAAAAAAAAAAAA 2417
Db 79 ATAAACCCCAAAAAAAAAAAAAAAAAA 54

RESULT 4

US-10-066-543-1610/c
; Sequence 1610, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jianguan
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1610
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 87, 252, 284, 315, 391
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-1610

Query Match 1.1%; Score 26; DB 14; Length 395;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 ATAAACCCCAAAAAAAAAAAAAAAAAA 2417
Db 40 ATAAACCCCAAAAAAAAAAAAAAAAAA 15

RESULT 5

US-10-494-343-4
; Sequence 4, Application US/10494343
; Publication No. US20040248138A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; APPLICANT: Phan, Thuyvy
; TITLE OF INVENTION: HUMAN AGIOMOTIN-LIKE PROTEIN 1
; FILE REFERENCE: P80184
; CURRENT APPLICATION NUMBER: US/10/494,343
; CURRENT FILING DATE: 2004-04-30
; PRIOR APPLICATION NUMBER: US to be assigned
; PRIOR FILING DATE: to be assigned
; PRIOR APPLICATION NUMBER: PCT/US2002/035129
; PRIOR FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/334,773
; PRIOR FILING DATE: 2001-11-01
; NUMBER OF SEQ ID NOS: 870

us-10-028-384-7.Oligo.rnpb

Thu Dec 16 16:25:11 2004

```
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 4
; LENGTH: 3248
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-494-343-4

Query Match      1.1%; Score 26; DB 18; Length 3248;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2392 ATAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 3207 ATAAACCCCAAAAAAAAAAAAAAAAAAAAAA 3232

RESULT 6
US-10-242-535A-53154
; Sequence 53154, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 53154
; LENGTH: 153
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (33)..(33)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-53154

Query Match      1.0%; Score 25; DB 16; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 122 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 146

RESULT 7
US-10-085-783A-14946
; Sequence 53154, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28

Query Match      1.0%; Score 25; DB 16; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 122 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 146

RESULT 8
US-10-242-535A-14946
; Sequence 14946, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14946
; LENGTH: 293
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-14946

Query Match      1.0%; Score 25; DB 16; Length 293;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2393 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
Db 260 TAAACCCCAAAAAAAAAAAAAAAAAAAAAA 284

RESULT 9
US-10-085-783A-14946
; Sequence 14946, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liow, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
```


FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_171793C.1
US-10-425-115-78691

Query Match 1.0%; Score 25; DB 18; Length 488;
Best Local Similarity 100.0%; Pred.No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

yy 2392 ATAAACCCAAAAAAAAAAAAAAAAAAAA 2416
|||
ob 179 ATAAACCCAAAAAAAAAAAAAAAAAAAA 155

RESULT 14
US-09-814-353-2031
Sequence 2031, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lille, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07

```

; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2031
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..110
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-2031

Query Match          1.0%; Score 25; DB 10; Length 513;
Best Local Similarity 100.0%; Pred.No. 0.043;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2393 TAACCCCAAAAAAAAAAAAAAAAAAAAAA 2417
         |||||
Db       486 TAACCCCAAAAAAAAAAAAAAAAAAAAAA 510

RESULT 15
US-09-814-353-8377
; Sequence 8377, Application US/09814353
; Publication NO. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814.353
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 10:16:54 ; Search time 11093.6 Seconds
(without alignments)
7939.229 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctagcagaagtgtctgc.....cccaaaaaaaaaaaaaaa 2417

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 32822875 seqs, 18219865908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gss1:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 734 | 30.4 | 810 | 2 | BF502026 | BF502026 AT17637.5 |
| 2 | 691 | 28.2 | 681 | 7 | CK659033 | CK659033 LP16462.5 |
| 3 | 664 | 27.5 | 701 | 2 | BF503420 | BF503420 AT19406.5 |
| 4 | 658 | 27.2 | 666 | 4 | BI357074 | BI357074 RE4325.5 |
| 5 | 658 | 27.2 | 668 | 4 | BI227902 | BI227902 RE25288.5 |
| 6 | 657 | 27.2 | 715 | 7 | CK657451 | CK657451 LP23759.5 |
| 7 | 652 | 27.0 | 652 | 4 | BI171940 | BI171940 RE13841.5 |
| 8 | 644 | 26.6 | 644 | 4 | BG636414 | BG636414 SD14123.5 |
| 9 | 644 | 26.6 | 654 | 4 | BI374334 | BI374334 RE61893.5 |
| 10 | 641 | 26.5 | 641 | 1 | AI257750 | AI257750 LP06212.5 |
| 11 | 641 | 26.5 | 679 | 7 | CK657649 | CK657649 LP24119.5 |
| 12 | 639 | 26.4 | 690 | 4 | BG640860 | BG640860 SD12024.5 |
| 13 | 635 | 26.3 | 642 | 4 | BI173041 | BI173041 RE15774.5 |
| 14 | 623 | 25.8 | 635 | 1 | AI135629 | AI135629 GH13452.5 |
| 15 | 615 | 25.4 | 655 | 2 | BF496296 | BF496296 AT10050.5 |
| 16 | 615 | 25.4 | 653 | 4 | BG641172 | BG641172 SD12448.5 |
| 17 | 612 | 25.3 | 615 | 4 | BI364613 | BI364613 RE49773.5 |
| 18 | 607 | 25.1 | 655 | 4 | BI484774 | BI484774 RE67926.5 |
| 19 | 600 | 24.8 | 600 | 4 | BI163592 | BI163592 RE03026.5 |
| 20 | 588 | 24.3 | 630 | 1 | AI295381 | AI295381 LP08987.5 |
| 21 | 586 | 24.2 | 616 | 2 | BF504378 | BF504378 AT05729.5 |
| 22 | 585 | 24.2 | 585 | 1 | AI133998 | AI133998 GH1327.5 |
| 23 | 576 | 23.8 | 585 | 4 | BI242003 | BI242003 RE39217.5 |
| 24 | 567 | 23.5 | 649 | 4 | BI162865 | BI162865 RE02124.5 |

| | | | | | | |
|----|-----|------|-----|---|----------|--------------------|
| 25 | 566 | 23.4 | 598 | 4 | BI170396 | BI170396 RE11825.5 |
| 26 | 564 | 23.3 | 658 | 4 | BI374189 | BI374189 RE61718.5 |
| 27 | 551 | 22.8 | 564 | 1 | AI542327 | AI542327 SD08615.5 |
| 28 | 549 | 22.7 | 735 | 1 | AA949890 | AA949890 LD29946.5 |
| 29 | 548 | 22.7 | 555 | 4 | BI370013 | BI370013 RE56375.5 |
| 30 | 538 | 22.3 | 682 | 7 | CK659064 | CK659064 LP16523.5 |
| 31 | 535 | 22.1 | 646 | 4 | BI170481 | BI170481 RE11944.5 |
| 32 | 529 | 21.9 | 554 | 4 | BI167839 | BI167839 RE08321.5 |
| 33 | 524 | 21.7 | 564 | 4 | BI635692 | BI635692 SD17119.5 |
| 34 | 511 | 21.1 | 525 | 2 | BF502679 | BF502679 AT18477.5 |
| 35 | 507 | 21.0 | 527 | 4 | BI586931 | BI586931 RH27159.5 |
| 36 | 507 | 21.0 | 539 | 1 | AI107144 | AI107144 GH06456.5 |
| 37 | 502 | 20.8 | 539 | 4 | BI373819 | BI373819 RE61243.5 |
| 38 | 495 | 20.5 | 495 | 1 | AI533215 | AI533215 SD04365.5 |
| 39 | 492 | 20.4 | 502 | 4 | BI237763 | BI237763 RE34052.5 |
| 40 | 480 | 19.9 | 480 | 1 | AI514952 | AI514952 LD46592.5 |
| 41 | 464 | 19.2 | 464 | 1 | AI297771 | AI297771 LP12143.5 |
| 42 | 458 | 18.9 | 478 | 1 | AI533035 | AI533035 SD04713.5 |
| 43 | 454 | 18.8 | 499 | 4 | BI173583 | BI173583 RE16531.5 |
| 44 | 441 | 18.2 | 445 | 4 | BI579570 | BI579570 RE73627.5 |
| 45 | 441 | 18.2 | 467 | 4 | BI364972 | BI364972 RE50181.5 |

ALIGNMENTS

RESULT 1
BF502026
LOCUS
DEFINITION
BF502026 810 bp mRNA linear EST 19-APR-2001
AT17637.Sprime AT Drosophila melanogaster adult testes P0877
Drosophila melanogaster cDNA clone AT17637 5 similar to Osetst3:
FBan0007748 'enzyme' located on: 3R 96B16-96B17: 04/09/2001, mRNA
sequence.
ACCESSION
BF502026
VERSION
BF502026.2 GI:13695997
KEYWORDS
EST
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Drosophila melanogaster
REFERENCE
1 (bases 1 to 810)
AUTHORS
Stapleton,M., Brockstein,P., Hong,L., Agbavani,A., Baxter,E.,
Berman,B., Carlson,J., Champagne,M., Chavez,C., Chew,M., Dorsett,V.,
Farfan,D., Faise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,
Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nuncio,J.,
Pacib,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,
Lewis,J.E., Celniker,S. and Rubin,G.M.
BDGP/HMI AT Drosophila EST Project
Unpublished (2000)
On Dec 6, 2000 this sequence version replaced gi:11585327.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003569; arm:X [20544566.20835979]
estimated-cyto:19E4-19F1: 04/09/2001 hit P element 1(3)j2D9:
1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
PCR: 02/10/2001
Plate: AT176 row: D column: 1
High quality sequence stop: 713.
Location/Qualifiers
1. 810
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT17637"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
AT.121-AT.319: DHS-alpha TonA"

/clone lib="AT Drosophila melanogaster adult testes POTB7"
/note="Organ: ADULT testes; Vector: POTB7; Site 1: EcoRI;
Site 2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into POTB7. Plasmid cDNA library."

ORIGIN

| | | | | |
|-----------------------|--------|-----------------|---------------|-------------------|
| Query Match | 30.4%; | Score 734; | DB 2; | Length 810; |
| Best Local Similarity | 99.9%; | Pred. No. 0; | | |
| Matches | 784; | Conservative 0; | Mismatches 1; | Indels 0; Gaps 0; |

| | | | |
|----|-----|--|-----|
| Qy | 23 | GCATTTTCAGATCGGTTTAAATTTTCGAGTTACTGGCTGGAATTTGGACATGAATCGGACG | 82 |
| Db | 1 | GCATTTTCAGATCGGTTTAAATTTTCGAGTTACTGGCTGGAATTTGGACATGAATCGGACG | 60 |
| Qy | 83 | CCGAAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCTTAATCACTTCGCAATCCTG | 142 |
| Db | 61 | CCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCTTAATCACTTCGCAATCCTG | 120 |
| Qy | 143 | CTAATCGCTGCTGCGGAGTTTCTCTCGCTCTTTCGCGTCTATCGCTTCAGCTTCGAGTCG | 202 |
| Db | 121 | CTAATCGCTGCTGCGGAGTTTCTCTCGCTCTTTCGCGTCTATCGCTTCAGCTTCGAGTCG | 180 |
| Qy | 203 | ATTATCCATGATGTTTGAATCGCTGCTCAACTACCGGCGCACCGCTTACATGCTGCAGAAAT | 262 |
| Db | 181 | ATTATCCATGATGTTTGAATCGCTGCTCAACTACCGGCGCACCGCTTACATGCTGCAGAAAT | 240 |
| Qy | 263 | GGTTGGTACAATCTTCTCAACTGTTTTCGAGCGGCGCATGTTATCGCTTCGCGAGGATT | 322 |
| Db | 241 | GGTTGGTACAATCTTCTCAACTGTTTTCGAGCGGCGCATGTTATCGCTTCGCGAGGATT | 300 |
| Qy | 323 | GTGGCGGTACCGTCTATCCCGGCTCATGATACGTCGCGGCGCAATCCATTGGCTGCTG | 382 |
| Db | 301 | GTGGCGGTACCGTCTATCCCGGCTCATGATACGTCGCGGCGCAATCCATTGGCTGCTG | 360 |
| Qy | 383 | CACGTACTCAACATACCGGCTCCATTTTCGTTGACATCTGGTGTTCGCGCGCATCTTC | 442 |
| Db | 361 | CACGTACTCAACATACCGGCTCCATTTTCGTTGACATCTGGTGTTCGCGCGCATCTTC | 420 |
| Qy | 443 | AGTGGCTGACCTCCATCTCCACCTACTGCTGACCAAGGAGCTGTGGTCCGCGGCGCC | 502 |
| Db | 421 | AGTGGCTGACCTCCATCTCCACCTACTGCTGACCAAGGAGCTGTGGTCCGCGGCGCC | 480 |
| Qy | 503 | GCCCTCTTCGCGCGCAGCTTCATCGCCATCGTGGCTGCTGCTGCTGCTGCTGCTGCTG | 562 |
| Db | 481 | GCCCTCTTCGCGCGCAGCTTCATCGCCATCGTGGCTGCTGCTGCTGCTGCTGCTGCTG | 540 |
| Qy | 563 | GGATCGTACGATACGAGGCGATTTGCCATATTGCGCCCTGCGATTCACCTACTTCTGTTG | 622 |
| Db | 541 | GGATCGTACGATACGAGGCGATTTGCCATATTGCGCCCTGCGATTCACCTACTTCTGTTG | 600 |
| Qy | 623 | GTGGCTCAGTGAAGATCTGATCCGTTCTGTTGCTGGCGCGCAGCTTGTCTTCTACTTC | 682 |
| Db | 601 | GTGGCTCAGTGAAGATCTGATCCGTTCTGTTGCTGGCGCGCAGCTTGTCTTCTACTTC | 660 |
| Qy | 683 | TACATGGTGTCCGCTGGGTTGGCTAGCTGTTTCATCATCAACCTGATACCCCTGCACTTC | 742 |
| Db | 661 | TACATGGTGTCCGCTGGGTTGGCTAGCTGTTTCATCATCAACCTGATACCCCTGCACTTC | 720 |
| Qy | 743 | TTCGTACTGCTCATTTATGGGAGGATCTCGCGCGCTGCTGACGACTACAGCACTTC | 802 |
| Db | 721 | TTCGTACTGCTCATTTATGGGAGGATCTCGCGCGCTGCTGACGACTACAGCACTTC | 780 |
| Qy | 803 | TACAT 807 | |
| Db | 781 | TACAT 785 | |

RESULT 2
CK659033
LOCUS

681 bp mRNA linear EST 30-JAN-2004

DEFINITION LP16462.Sprime LP Drosophila melanogaster larval-early pupal pOT2
Drosophila melanogaster cDNA clone LP16462 5, mRNA sequence.
CK659033
VERSION CK659033.1 GI:41402558
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 681)
AUTHORS Harvey D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S., and Rubin, G.M.
BDGP/HMIR Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd., Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: LP.164 row: F column: 2
High quality sequence stop: 670.
FEATURES
1..681
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP16462"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."
ORIGIN

| | | | | |
|-----------------------|---------|-----------------|---------------|-------------------|
| Query Match | 28.2%; | Score 681; | DB 7; | Length 681; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches | 681; | Conservative 0; | Mismatches 0; | Indels 0; Gaps 0; |

| | | | |
|----|-----|--|-----|
| Qy | 16 | TGTCGTTCATTTTCAGATCGGTTTAAATTTTCGAGTTACTGGCTGGAATTTGGACATGA | 75 |
| Db | 1 | TGTCGTTCATTTTCAGATCGGTTTAAATTTTCGAGTTACTGGCTGGAATTTGGACATGA | 60 |
| Qy | 76 | TCGAGCGCGCAAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCTTAATCACTTCGC | 135 |
| Db | 61 | TCGAGCGCGCAAGATGCTGAACAGCAAGTGGCTGCTACAGCAGCTTAATCACTTCGC | 120 |
| Qy | 136 | CATCCTGCTAAATCGCTGCGGATTTTCTCTCGCTTTCGCTTCGCTTCATCGCTT | 195 |
| Db | 121 | CATCCTGCTAAATCGCTGCGGATTTTCTCTCGCTTTCGCTTCGCTTCATCGCTT | 180 |
| Qy | 196 | CGAGTCGATTTATCCATGATTTGATCCGTTGTTCACTACCGGCGCACCGCTACATGCT | 255 |
| Db | 181 | CGAGTCGATTTATCCATGATTTGATCCGTTGTTCACTACCGGCGCACCGCTACATGCT | 240 |
| Qy | 256 | GCAGATGTTGGTGAACACTTCTCACTGGTTGACGAGCGCGCATGTTATCGCTCGG | 315 |
| Db | 241 | GCAGATGTTGGTGAACACTTCTCACTGGTTGACGAGCGCGCATGTTATCGCTCGG | 300 |
| Qy | 316 | CAGATTTGGGCGGTACCGTCTATCCCGGCGCTGATGATTACGTCGCGCGGAATCCATTG | 375 |
| Db | 301 | CAGATTTGGGCGGTACCGTCTATCCCGGCGCTGATGATTACGTCGCGCGGAATCCATTG | 360 |
| Qy | 376 | GCTGCTGACGCTACTCAACATACCGGTCCATATTCTGTCGATCTGCTTCTGGGCGCC | 435 |
| Db | 361 | GCTGCTGACGCTACTCAACATACCGGTCCATATTCTGTCGATCTGCTTCTGGGCGCC | 420 |
| Qy | 436 | GATCTTCAGTGGCGGTGACCTCCATCTCCACCTACCTGCTGCTACCAAGAGCTGTGTCGC | 495 |

Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
Rubin, G.M.
BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic AB003750: arm:3R [20671983, 20899380]
estimated-cyto:96B10-96C1: 05/13/2001 hit P element 1(3)J2D9:
1(3)J2D9 A026308 inserted at base 292 5', end of P element Inverse
PCR: 05/13/2001
Plate: RE:434 row: C column: 1
High quality sequence stop: 563.

FEATURES
source
1..666
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE43425"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
p1c-1"
/note="Organ: embryo; Vector: p1c1; Site 1: XhoI; Site 2:
BamHI. Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 27.2%; Score 658; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCTAAGCGAAGATGTCGTCGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
Db 9 TCTAAGCGAAGATGTCGTCGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 68
Qy 61 GAATTGGGACATGAATCGGACCGCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 120
Db 69 GAATTGGGACATGAATCGGACCGCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 128
Qy 121 CCTAATCACCTTCGCCATCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 180
Db 129 CCTAATCACCTTCGCCATCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 188
Qy 181 CGCGTCATCCGTTCCGAGTCGATTCATTCAGATGTTGATCCGTTCACTACCGGCG 240
Db 189 CGCGTCATCCGTTCCGAGTCGATTCATTCAGATGTTGATCCGTTCACTACCGGCG 248
Qy 241 CACGCGCTACATGCTGAGAAATGTTGGTACAACTTCTCAACTGTTTCGACGAGCGC 300
Db 249 CACGCGCTACATGCTGAGAAATGTTGGTACAACTTCTCAACTGTTTCGACGAGCGC 308
Qy 301 ATGTATTCGCTCGGCGGATTTGGCGGTCATCGCTATCCCGGCTGATGATTCGTC 360
Db 309 ATGTATTCGCTCGGCGGATTTGGCGGTCATCGCTATCCCGGCTGATGATTCGTC 368
Qy 361 CGGCGGAATCCATGGCTGCTGACGCTACCTCAACATACCGGTCATATTCGTCGATCTG 420
Db 369 CGGCGGAATCCATGGCTGCTGACGCTACCTCAACATACCGGTCATATTCGTCGATCTG 428
Qy 421 CGTGTTCGTCGCGCGATCTTCAGTGCCCTGACCTCCATCTCCACCTACCTGCTGACCA 480
Db 429 CGTGTTCGTCGCGCGATCTTCAGTGCCCTGACCTCCATCTCCACCTACCTGCTGACCA 488
Qy 481 GGAGCTGTGTCGCGCGGCGCGGCTCTTCGCGCGGAGCTTCATGCCATCGTGGCTGG 540
Db 489 GGAGCTGTGTCGCGCGGCGCGGCTCTTCGCGCGGAGCTTCATGCCATCGTGGCTGG 548

Qy 541 CTACATCAGTAGTGGTGGCTGGATCGATACGAGGCAATTCGATATTCGCTC 600
Db 549 CTACATCAGTAGTGGTGGCTGGATCGATACGAGGCAATTCGATATTCGCTC 608
Qy 601 GCAGTTCACTACTTCTGCTGGTGGCTCATGAGACTGATCGTGTCTGCTCG 658
Db 609 GCAGTTCACTACTTCTGCTGGTGGCTCATGAGACTGATCGTGTCTGCTCG 666

RESULT 5
LOCUS BI227902 668 bp mRNA linear EST 11-JUL-2001
DEFINITION RE25288.Sprime RE Drosophila melanogaster normalized Embryo p1c-1
Drosophila melanogaster cDNA clone RE25288 5 similar to Osetctt3:
Pfan0007748 'enzyme' located on: 3R 96B16-96B17;: 04/12/2001, mRNA
sequence.
ACCESSION BI227902 GI:14695166
VERSION BI227902.1
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 668)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Miera, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic AB003750: arm:3R [20671983, 20899380]
estimated-cyto:96B10-96C1: 04/12/2001 hit P element 1(3)J2D9:
1(3)J2D9 A026308 inserted at base 292 5', end of P element Inverse
PCR: 03/20/2001
Plate: RE:252 row: H column: 4
High quality sequence stop: 520.

FEATURES
source
1..668
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE25288"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
p1c-1"
/note="Organ: embryo; Vector: p1c1; Site 1: XhoI; Site 2:
BamHI. Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 27.2%; Score 658; DB 4; Length 668;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCTAAGCGAAGATGTCGTCGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
Db 11 TCTAAGCGAAGATGTCGTCGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 70
Qy 61 GAATTGGGACATGAATCGGACCGCGAGATGCTGAACAGCAAGTGGCTGCTACAGCAG 120

REFERENCE 1 (bases 1 to 652)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Chan, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guerin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
 TITLE BDGP/HMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R (20671983, 20899380)
 estimated-cyto: 96B10-96C1: 04/06/2001 hit P element 1(3)J2D9:
 1(3)J2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 04/06/2001
 Plate: RE.138 row: D column: 5
 High quality sequence stop: 590.
 Location/Qualifiers
 1..652
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE13841"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_hosts="DHS-alpha Tora"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pflc-1"
 /note="Organ: embryo; Vector: pELc1; Site: 1: XhoI; Site: 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
 ORIGIN
 Query Match 27.0%; Score 652; DB 4; Length 652;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 652; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 AATGTCGTCGATTCAGATCGGTTATATTTTCAGTTACTGCTGGATTCGGACA 71
 DB 1 AATGTCGTCGATTCAGATCGGTTATATTTTCAGTTACTGCTGGATTCGGACA 60
 QY 72 TGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTTAATCACT 131
 DB 61 TGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTTAATCACT 120
 QY 132 TCGCATCTCTGCTAATCGCTGCTGGCGGATTTTCCTCTGCTGCTCTTCGCGTCATCC 191
 DB 121 TCGCATCTCTGCTAATCGCTGCTGGCGGATTTTCCTCTGCTGCTCTTCGCGTCATCC 180
 QY 192 GTTTCGAGTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGGCGACCGCTACA 251
 DB 181 GTTTCGAGTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGGCGACCGCTACA 240
 QY 252 TGGTGCAGATGTTGGTGAACAATCTCTCACTGCTGCAAGCGGCGCGATTCGCTGC 311
 DB 241 TGGTGCAGATGTTGGTGAACAATCTCTCACTGCTGCAAGCGGCGCGATTCGCTGC 300
 QY 312 TCGGACGAGATTGGCGGCGTACCGTCTATCCCGGCTGATGATTACGTCGCGGAATCC 371
 DB 301 TCGGACGAGATTGGCGGCGTACCGTCTATCCCGGCTGATGATTACGTCGCGGAATCC 360
 QY 372 ATTGGCTCTCAGTACTCAACATACCGGCTCATATTCGTGACATCTGCGTTCCTGG 431
 DB 361 ATTGGCTCTCAGTACTCAACATACCGGCTCATATTCGTGACATCTGCGTTCCTGG 420
 QY 432 CGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACTGCTGCTACCAAGGAGCTGGT 491
 DB 421 CGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACTGCTGCTACCAAGGAGCTGGT 480

QY 492 CGCGGGGGCGGGCGCTCTTCGCGCGAGCTTCATCGGCATCGTGGCTGCTACATCAGTA 551
 DB 481 CGCGGGGGCGGGCGCTCTTCGCGCGAGCTTCATCGGCATCGTGGCTGCTACATCAGTA 540
 QY 552 GGTCTGGTGGCTGGATCGTACGATTAACAGAGGCAATTCGCATATTCGCGCTGCAGTTCACT 611
 DB 541 GGTCTGGTGGCTGGATCGTACGATTAACAGAGGCAATTCGCATATTCGCGCTGCAGTTCACT 600
 QY 612 ACTTCTCTGGTGGCTGGCTCAGTGAAGACTGATCGTGGTCTGGTGGCGGC 663
 DB 601 ACTTCTCTGGTGGCTGGCTCAGTGAAGACTGATCGTGGTCTGGTGGCGGC 652

RESULT 8
 LOCUS BG636414
 DEFINITION SD14123.Sprine SD Drosophila melanogaster Schneider L2 cell culture
 POT2 Drosophila melanogaster cDNA clone SD14123 5 similar to
 Oet8t3: FBan000748 'enzyme' located on: 3R 96B16-96B17;
 04/13/2001, mRNA sequence.
 ACCESSION BG636414
 VERSION BG636414.1 GI:13763951
 KEYWORDS EST.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 644)
 AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Teang, G., Lewis, S. and Rubin, G.M.
 TITLE BDGP/HMI Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R (20671983, 20899380)
 estimated-cyto: 96B10-96C1: 04/13/2001 hit P element 1(3)J2D9:
 1(3)J2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 03/20/2001
 Plate: SD.141 row: B column: 11
 High quality sequence stop: 612.
 Location/Qualifiers
 1..644
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="SD14123"
 /lab_hosts="SD Drosophila melanogaster Schneider L2 cell culture POT2"
 /note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2.
 Plasmid cDNA library."

Query Match 26.6%; Score 644; DB 4; Length 644;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCTAAGCGAAGATGTCGTTGCATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
 DB 1 TCTAAGCGAAGATGTCGTTGCATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
 QY 61 GAATTGGACATGATCGGACCGCGAGAGCTGTAACAGCAAGTGGCTGCTACAGCAG 120
 DB 61 GAATTGGACATGATCGGACCGCGAGAGCTGTAACAGCAAGTGGCTGCTACAGCAG 120
 QY 121 CCTAATACCTTCGCTCAATCTGCTAATCGCTGCTGGCGGATTTTCCTCTCGCTCTT 180

```

121 CCTAATCACTTCGCGCATCTGCTAATCGCTGGCTGCGCGAATTTCTCTCGCCTTT 180
181 CGCGCTCATCGTTTTCGATTCGATTCATCAGTTCATCGTGGTTCACTACCGGGC 240
181 CGCGCTCATCGTTTTCGATTCGATTCATCAGTTCATCGTGGTTCACTACCGGGC 240
241 CACCGCTACATGTTGTCAGAAATGTTGTTGTAACAATTTCTCAACTTCTCACTGGTTGAGAGCGCGC 300
241 CACCGCTACATGTTGTCAGAAATGTTGTTGTAACAATTTCTCAACTTCTCACTGGTTGAGAGCGCGC 300
301 ATGGTATCGCTTCGCGAGATGTTGGCGGTTACCGTTCATCCCGGCTGATGATGTC 360
301 ATGGTATCGCTTCGCGAGATGTTGGCGGTTACCGTTCATCCCGGCTGATGATGTC 360
361 CGCGGGAATCCATTCGCTGCTGCACTACTCAACATACCGGTCCATATTCGTGACATCTG 420
361 CGCGGGAATCCATTCGCTGCTGCACTACTCAACATACCGGTCCATATTCGTGACATCTG 420
421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCGACAA 480
421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTCGACAA 480
481 GGAGCTGTGGTTCGCGCGCGCGCTCTTCGCGCGAGCTTCATCCGCTGCGCTGG 540
481 GGAGCTGTGGTTCGCGCGCGCGCTCTTCGCGCGAGCTTCATCCGCTGCGCTGG 540
541 CTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
541 CTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
601 GCAGTTCACCTACTTCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 644
601 GCAGTTCACCTACTTCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 644

```

```

RESULT 9
BI374334
LOCUS
DEFINITION
  654 bp mRNA linear EST 01-AUG-2001
  RE1893: Spine RE Drosophila melanogaster normalized Embryo pf1c-1
  Drosophila melanogaster cDNA clone RE1893 5 similar to OstStt3:
  FBan0007748 GO: 0101g0saccharide transferase (GO:0004576); enzyme
  (GO:0003824); located on: 3R 96B16-96B17; 05/16/2001, mRNA
  sequence.
ACCESSION
  BI374334
  BI374334.1 GI:15070362
KEYWORDS
  EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 654)
REFERENCE
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champagne,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzales,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Pakagas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AB003750: arm3R [20671983,20899380]
  estimated-cyto:96B10-96C1: 05/16/2001 hit P element 1(3)J2D9:
  1(3)J2D9 AQ026308 inserted at base 292 5', end of P element Inverse
  PCR: 05/16/2001
  Plate: RE.618 row: H column: 9
  High quality sequence stop: 595.

```

```

FEATURES
  source
  1..654
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE1893"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DHS-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pf1c-1"
  /notes="Organ: embryo; Vector: pf1c1; Site 1: XhoI; Site 2:
  BamHI; library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."
  ORIGIN
  Query Match 26.6%; Score 644; DB 4; Length 654;
  Best Local Similarity 100.0%; Pred. No. 0;
  Matches 644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1 TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 60
  Db 11 TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTATATATTTTCGAGTTACTGGCTG 70
  QY 61 GAATTGGGACATGAATTCGGACCCGGAAGATGCTGAACAGCAAGTGGTGGCTACAGCAG 120
  Db 71 GAATTGGGACATGAATTCGGACCCGGAAGATGCTGAACAGCAAGTGGTGGCTACAGCAG 130
  QY 121 CTTAATCAGTTTGGCCATCTCTTAATCGCCTGGCTGGCCGGAATTTCTCTCGCCTCTT 180
  Db 131 CTTAATCAGTTTGGCCATCTCTTAATCGCCTGGCTGGCCGGAATTTCTCTCGCCTCTT 190
  QY 181 CGCGGTATCCGTTTTCGAGTTCGATTCATCGATGTTGATCGGTGGTTCACATACCGGGC 240
  Db 191 CGCGGTATCCGTTTTCGAGTTCGATTCATCGATGTTGATCGGTGGTTCACATACCGGGC 250
  QY 241 CACCGCTACATGTTGTCAGAAATGTTGTAACAATTTCTCAACTGTTTCGAGAGCGCGC 300
  Db 251 CACCGCTACATGTTGTCAGAAATGTTGTAACAATTTCTCAACTGTTTCGAGAGCGCGC 310
  QY 301 ATGGTATCCGCTCGCGAGATGTTGGCGGTTACCGTTCATCCCGGCTGATGATGATGTC 360
  Db 311 ATGGTATCCGCTCGCGAGATGTTGGCGGTTACCGTTCATCCCGGCTGATGATGATGTC 370
  QY 361 CGCGGGAATCCATTCGCTGCTGCACTACTCAACATACCGGTTCATATTCGTGACATCTG 420
  Db 371 CGCGGGAATCCATTCGCTGCTGCACTACTCAACATACCGGTTCATATTCGTGACATCTG 430
  QY 421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
  Db 431 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 490
  QY 481 GGAGCTGTGGTTCGCGAGATGTTGGCGGTTACCGTTCATCCCGGCTGATGATGATGTC 540
  Db 491 GGAGCTGTGGTTCGCGAGATGTTGGCGGTTACCGTTCATCCCGGCTGATGATGATGTC 550
  QY 541 CTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
  Db 551 CTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 610
  QY 601 GCAGTTCACCTACTTCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 644
  Db 611 GCAGTTCACCTACTTCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 654

```

```

RESULT 10
A1257750
LOCUS
DEFINITION
  641 bp mRNA linear EST 19-APR-2001
  LP06212: Spine LP Drosophila melanogaster larval-early pupal poT2
  Drosophila melanogaster cDNA clone LP06212 5prime, mRNA sequence.
ACCESSION
  A1257750
  A1257750.1 GI:3865275

```

| | | | | |
|---------------------------|--|---|---|-----------------|
| KEYWORDS | EST. | 497 | GGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCTGCTACATCAGTAGGTCG | 556 |
| SOURCE | Drosophila melanogaster (fruit fly) | 481 | GGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCTGCTACATCAGTAGGTCG | 540 |
| ORGANISM | Drosophila melanogaster | 557 | GTGGCTGGATTCGATACGATTAACGAGGCGATTGCATATTGCGCCCTCAGTTCACTTAC | 616 |
| | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | 541 | GTGGCTGGATTCGATACGATTAACGAGGCGATTGCATATTGCGCCCTCAGTTCACTTAC | 600 |
| REFERENCE | 1 (bases 1 to 641) | 617 | CTGTGGTGGCTCAGTGAAGACTGATCGGTCTTGGTC | 657 |
| AUTHORS | Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M. | 601 | CTGTGGTGGCTCAGTGAAGACTGATCGGTCTTGGTC | 641 |
| TITLE | BDGP/HMI Drosophila EST Project | | | |
| JOURNAL | Unpublished (2001) | | | |
| COMMENT | Contact: Stapleton, M. | | | |
| | BDGP | | | |
| FEATURES | source | | | |
| | 1..641 | | | |
| | /organism="Drosophila melanogaster" | | | |
| | /mol_type="mRNA" | | | |
| | /db_xref="taxon:7227" | | | |
| | /clone="LP06212" | | | |
| | /sex="male and female" | | | |
| | /dev_stage="larvae-pupae" | | | |
| | /lab_host="DH5-alpha" | | | |
| | /clone_lib="LP Drosophila melanogaster larval-early pupal pot2" | | | |
| | /note="Organ: whole body; Vector: pot2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library." | | | |
| ORIGIN | | | | |
| Query Match | 26.5%; Score 641; DB 1; Length 641; | | | |
| Best Local Similarity | 100.0%; Pred. No. 0; | | | |
| Matches 641; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | | | |
| Qy | 17 | GTGCTGATTCAGATCGGTTATATTTTCGATTCTGCTGATCGGTTGGAATGGACATGAAT | 76 | |
| Db | 1 | GTGCTGATTCAGATCGGTTATATTTTCGATTCTGCTGATCGGTTGGAATGGACATGAAT | 60 | |
| Qy | 77 | CGGACGCGGAGATGCTGACACAGAGTGGCTGCTACAGACGCTATACCTTCGCG | 136 | |
| Db | 61 | CGGACGCGGAGATGCTGACACAGAGTGGCTGCTACAGACGCTATACCTTCGCG | 120 | |
| Qy | 137 | ATCCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTTCGCGCTCATCGGTTTC | 196 | |
| Db | 121 | ATCCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGCTCTTCGCGCTCATCGGTTTC | 180 | |
| Qy | 197 | GAGTCGATTCATGATGTTTGTATCGTGTTCACCTACCGGCGACCGCTACATGGTG | 256 | |
| Db | 181 | GAGTCGATTCATGATGTTTGTATCGTGTTCACCTACCGGCGACCGCTACATGGTG | 240 | |
| Qy | 257 | CAGAATGGTTGGTACAACTTCCTCAACTGCTTCGACGAGCGCGCATGGTATCGGTCGCG | 316 | |
| Db | 241 | CAGAATGGTTGGTACAACTTCCTCAACTGCTTCGACGAGCGCGCATGGTATCGGTCGCG | 300 | |
| Qy | 317 | AGGATTCGGCGGTACCGTCTATCCGCGCTGATGATTCGTCGCGCGGATTCATGG | 376 | |
| Db | 301 | AGGATTCGGCGGTACCGTCTATCCGCGCTGATGATTCGTCGCGCGGATTCATGG | 360 | |
| Qy | 377 | CTGCTGACGCTACTCAACATACCGGTCCATATTCGTGACATCTGCTGTTCCTGGCGCG | 436 | |
| Db | 361 | CTGCTGACGCTACTCAACATACCGGTCCATATTCGTGACATCTGCTGTTCCTGGCGCG | 420 | |
| Qy | 437 | ATCTTCAGTGGCTGACCTCATCTCACCCTACTCTGACCAAGAGCTGTGGTCGCG | 496 | |
| Db | 421 | ATCTTCAGTGGCTGACCTCATCTCACCCTACTCTGACCAAGAGCTGTGGTCGCG | 480 | |
| | | | | |
| RESULT 11 | | | | |
| LOCUS | CK657649 | 679 bp | mRNA | linear |
| DEFINITION | LP24119.Sprime LP Drosophila melanogaster larval-early pupal pot2 | | | EST 30-JAN-2004 |
| | Drosophila melanogaster cDNA clone LP24119 5, mRNA sequence. | | | |
| ACCESSION | CK657649 | | | |
| VERSION | CK657649.1 | GI:41401174 | | |
| KEYWORDS | EST. | | | |
| SOURCE | Drosophila melanogaster (fruit fly) | | | |
| ORGANISM | Drosophila melanogaster | | | |
| | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | |
| REFERENCE | 1 (bases 1 to 679) | | | |
| AUTHORS | Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M. | | | |
| TITLE | BDGP/HMI Drosophila EST Project | | | |
| JOURNAL | Unpublished (2001) | | | |
| COMMENT | Contact: Stapleton, M. | | | |
| | BDGP | | | |
| | Lawrence Berkeley National Lab | | | |
| | One Cyclotron Rd, Berkeley, CA 94720, USA | | | |
| | Fax: 510 486 6798 | | | |
| | Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu | | | |
| | hit p element sequence l(3)j2D9 | | | |
| | Plate: 62 row: A column: 12 | | | |
| | High quality sequence stop: 462. | | | |
| | Location/Qualifiers | | | |
| | 1..679 | | | |
| | /organism="Drosophila melanogaster" | | | |
| | /mol_type="mRNA" | | | |
| | /db_xref="taxon:7227" | | | |
| | /clone="LP24119" | | | |
| | /sex="male and female" | | | |
| | /dev_stage="larvae-pupae" | | | |
| | /lab_host="DH5-alpha" | | | |
| | /clone_lib="LP Drosophila melanogaster larval-early pupal pot2" | | | |
| | /note="Organ: whole body; Vector: pot2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library." | | | |
| ORIGIN | | | | |
| Query Match | 26.5%; Score 641; DB 7; Length 679; | | | |
| Best Local Similarity | 100.0%; Pred. No. 0; | | | |
| Matches 641; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | | | |
| Qy | 56 | GGCTGGAATGGGACATGAATCGGACGCGCGAAGATCTGAACAGAGTGGCTGCTAC | 115 | |
| Db | 39 | GGCTGGAATGGGACATGAATCGGACGCGCGAAGATCTGAACAGAGTGGCTGCTAC | 98 | |
| Qy | 116 | AGGACCTATACCTTCGCGCTTCCTGCTTAATCGCTGGCTGCGCGATTTCTCTCG | 175 | |
| Db | 99 | AGGACCTATACCTTCGCGCTTCCTGCTTAATCGCTGGCTGCGCGATTTCTCTCG | 158 | |
| Qy | 176 | CTCTTCGCGCTTCCTGCGCTTCCTGCTTAATCGCTGGCTGCGCGATTTCTCTCG | 235 | |
| Db | 159 | CTCTTCGCGCTTCCTGCGCTTCCTGCTTAATCGCTGGCTGCGCGATTTCTCTCG | 218 | |
| Qy | 236 | CGGCGCACCGCTACATGGTGGTGGTACAACTTCCTCAACTGGTTCGAG | 295 | |

Db 219 CGGGCCACCGCTCATGATGAGAAATGTTGGTCAAACTTCCCACTGGTTCGAGAG 278
 QY 296 CCGCATGATATCCGCTCGGAGGATTTGGGCGGTACCGTCTATCCGCGCTGATGATT 355
 Db 279 CGCGATGATATCCGCTCGGAGGATTTGGGCGGTACCGTCTATCCGCGCTGATGATT 338
 QY 356 AGCTCGGCGGGAATCCATTGCTGTCAGCTGCTCAACATACCGGTCCATATTCGAGAC 415
 Db 339 AGCTCGGCGGGAATCCATTGCTGTCAGCTGCTCAACATACCGGTCCATATTCGAGAC 338
 QY 416 ATCTCGGCTTCTCGGCGGATTTCTAGTGGCTGACCTCCATCTCCACCTACCTGCTG 475
 Db 399 ATCTCGGCTTCTCGGCGGATTTCTAGTGGCTGACCTCCATCTCCACCTACCTGCTG 458
 QY 476 ACCAAGAGCTGTTGCTCGGCGGCGCGGCTCTTCGCGCGGAGCTTCATCGCCATCGTG 535
 Db 459 ACCAAGAGCTGTTGCTCGGCGGCGCGGCTCTTCGCGCGGAGCTTCATCGCCATCGTG 518
 QY 536 CTGGGTATACAGTAGTGGTGGCTGATGATGATGATGATGATGATGATGATGATGATG 595
 Db 519 CTGGGTATACAGTAGTGGTGGCTGATGATGATGATGATGATGATGATGATGATGATG 578
 QY 596 GCGCTGACCTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 655
 Db 579 GCGCTGACCTTCACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
 QY 656 TCGGCGGCGAGCGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 696
 Db 639 TCGGCGGCGAGCGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 679

RESULT 12

EG640860 690 bp mRNA linear EST 23-APR-2001
 SD12024. Sprime SD Drosophila melanogaster Schneider L2 cell culture
 LOCUS Drosophila melanogaster cDNA clone SD12024 5 similar to
 Defn: OtsStt3; Fban007748 'enzyme' located on: 3R 96B16-96B17; ;
 04/13/2001, mRNA sequence.

EG640860
 EG640860.1 GI:13727286

EST.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 690)
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S., and Rubin, G.M.

BDGP/HMI Drosophila EST Project
 Unpublished (2001)
 Contact: Stapleton, M.
 BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AB003750; arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1: 04/13/2001 hit P element 1(3)j2D9:
 1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 03/20/2001

Plate: SD120 row: B column: 12
 High quality sequence stop: 535.

FEATURES

source

1..690
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="SD12024"
 /lab_host="DH5-alpha"
 /clone_lib="SD Drosophila melanogaster Schneider L2 cell
 culture pOT2"

/note="vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
 fractionated cDNAs were directly ligated into pOT2.
 Plasmid cDNA library."

ORIGIN
 Query Match 26.4%; Score 639; DB 4; Length 690;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 689; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 21 TTGCATTTTCAGATCGTTTATTAATTTTCGAGTTACTGCTGGAATTTGGACATGAATCGGA 80
 Db 1 TTGCATTTTCAGATCGTTTATTAATTTTCGAGTTACTGCTGGAATTTGGACATGAATCGGA 60
 QY 81 CGCCCAAGATGCTGAACAGCAAGCTGGCTGGCTACAGCAGCTTAATCACCTTCGCCATCC 140
 Db 61 CGCCCAAGATGCTGAACAGCAAGCTGGCTGGCTACAGCAGCTTAATCACCTTCGCCATCC 120
 QY 141 TCTTAATCGCTGGCTGGCGGATTTCTCTGCGCTCTTCGCGCTCATCGTTTCGAGT 200
 Db 121 TCTTAATCGCTGGCTGGCGGATTTCTCTGCGCTCTTCGCGCTCATCGTTTCGAGT 180
 QY 201 CGATTATCCATGAGTTTTCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 260
 Db 181 CGATTATCCATGAGTTTTCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 240
 QY 261 ATGGTTGGTACAACTTCTCAACTTTCGAGAGCGCGCATGGTATCCGCTCGCGAGGA 320
 Db 241 ATGGTTGGTACAACTTCTCAACTTTCGAGAGCGCGCATGGTATCCGCTCGCGAGGA 300
 QY 321 TTGTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTCGGCGGGAATCCATTGGCTGC 380
 Db 301 TTGTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTCGGCGGGAATCCATTGGCTGC 360
 QY 381 TGCAGTACTCAACATACCGCTTCATATTCGTCATCTGCTGGCTTCTCGCGCGCATCT 440
 Db 361 TGCAGTACTCAACATACCGCTTCATATTCGTCATCTGCTGGCTTCTCGCGCGCATCT 420
 QY 441 TCAGTGGCTGACCTCCCATCTCCACCTACCTGCTGACCAAGAGCTGCTGGTCCGCGGCG 500
 Db 421 TCAGTGGCTGACCTCCCATCTCCACCTACCTGCTGACCAAGAGCTGCTGGTCCGCGGCG 480
 QY 501 CCGGCTCTTTCGCGCGCATGTTTCATTCGCTTCGCTGGCTTCATCATAGTAGTTCGCTGG 560
 Db 481 CCGGCTCTTTCGCGCGCATGTTTCATTCGCTTCGCTGGCTTCATCATAGTAGTTCGCTGG 540
 QY 561 CTGATCTGATAGTAAAGAGCGCATTCGCATATTCGCGCTTCAGTTTCCACTTCTTCTCT 620
 Db 541 CTGATCTGATAGTAAAGAGCGCATTCGCATATTCGCGCTTCAGTTTCCACTTCTTCTCT 600
 QY 621 GGGTGGCTGATGAAAGCTGGATTCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 680
 Db 601 GGGTGGCTGATGAAAGCTGGATTCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 660
 QY 681 TCTCATATGCTTCGCGCTGGGCTGGCTACG 710
 Db 661 TCTCATATGCTTCGCGCTGGGCTGGCTACG 690

RESULT 13

BI173041

LOCUS

DEFINITION

ORGANISM

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM

ORGANISM


```
QY 189 TCCGTTTCGAGTCGATTATCCATGAGTTTGATCGTGGTCAACTACCGGGCCACCGCCT 248
Db 181 TCCGTTTCGAGTCGATTATCCATGAGTTTGATCGTGGTCAACTACCGGGCCACCGCCT 240
QY 249 ACATGGTGCAGAAATGGTGTGATACAACTTCTCACTGTTTCCGAGCGCGCATGTATC 308
Db 241 ACATGGTGCAGAAATGGTGTGATACAACTTCTCACTGTTTCCGAGCGCGCATGTATC 300
QY 309 CGCTCGGAGAAATGGTGGGGGTACCGCTATCCCGGCTGATGATTACGTCGGGGGAA 368
Db 301 CGCTCGGAGAAATGGTGGGGGTACCGCTATCCCGGCTGATGATTACGTCGGGGGAA 360
QY 369 TCCATTGGCTGCTGAGTACTCAACATACCGGTCCATATCGTGACATCGTGTTC 428
Db 361 TCCATTGGCTGCTGAGTACTCAACATACCGGTCCATATCGTGACATCGTGTTC 420
QY 429 TGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGT 488
Db 421 TGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGT 480
QY 489 GGTCCGGGGGCGCGGCTCTTCGCCCGGAGCTTCATCGGCATCGTGGCTGATATCA 548
Db 481 GGTCCGGGGGCGCGGCTCTTCGCCCGGAGCTTCATCGGCATCGTGGCTGATATCA 540
QY 549 GTAGTTCGGTGGCTGATCGTACGATACGAGGCGATTGCGCATATTCGCCCTCGAGTTCA 608
Db 541 GTAGTTCGGTGGCTGATCGTACGATACGAGGCGATTGCGCATATTCGCCCTCGAGTTCA 600
QY 609 CTTACTTCTCTGGTGGCTGCTCA 631
Db 601 CTTACTTCTCTGGTGGCTGCTCA 623
```

RESULT 15

BF496296

LOCUS

DEFINITION

BF496296 655 bp mRNA linear EST 19-APR-2001

Drosophila melanogaster adult testes pOTB7

Drosophila melanogaster cDNA clone AT10060 5 similar to OstStt3:

Fban0007748 'enzyme' located on: 3R 96B16-96B17; 04/08/2001, mRNA

sequence.

BF496296

BF496296.2 GI:13693786

EST.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophiliidae; Drosophila.

1 (bases 1 to 655)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,

Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,

Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N.,

Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J.,

Paclob, J., Paragas, V., Park, S., Phuanavong, S., Wan, K., Yu, C.,

Lewis, S.E., Celniker, S. and Rubin, G.M.

BDGP/HMI At Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11579597.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AE003750: arm:3R [20671983, 20899380]

estimated-cyto:96B10-96C1: 04/08/2001 hit P element 1(3)j2D9:

1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse

PCR: 02/09/2001

Plate: AT100 row: E column: 12

High quality sequence stop: 577.

Location/Qualifiers

1. .655

/organism="Drosophila melanogaster"

FEATURES

source

```
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT10060"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT10-AT120: DHS-alpha. Plates
AT121-AT319: DHS-alpha Tona"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/notes="Organ: ADULT testes; Vector: pOTB7; Site 1: EcORI;
Site 2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
```

ORIGIN

```
Query Match 25.4%; Score 615; DB 2; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.3e-309; Indels 0; Gaps 0;
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 CTGGAATTGGGACATGAATCGGACCGCCGAAAGATGCTGAACAGCAAGTGGTGGCTACAG 117
Db 1 CTGGAATTGGGACATGAATCGGACCGCCGAAAGATGCTGAACAGCAAGTGGTGGCTACAG 60
QY 118 CAGCCTTAATCACTTCGCGCATCTCTCTAATCGCCTGGCTGGCGGATTTCTCTCGCCT 177
Db 61 CAGCCTTAATCACTTCGCGCATCTCTCTAATCGCCTGGCTGGCGGATTTCTCTCGCCT 120
QY 178 CTTCCCGCTCATCCGTTTCGAGTCGATTCATCCATGAGTTTGATCGGTGGTTCAATACCG 237
Db 121 CTTCCCGCTCATCCGTTTCGAGTCGATTCATCCATGAGTTTGATCGGTGGTTCAATACCG 180
QY 238 GGCACACCGCTACATGGTGCAGATGGTGTGATCAACTTCTCACTGGTTCGAGGAGCG 297
Db 181 GGCACACCGCTACATGGTGCAGATGGTGTGATCAACTTCTCACTGGTTCGAGGAGCG 240
QY 298 CGCATGGTATCCGCTCGGACGATTTGGGGCGGTACCGCTCTATCCCGGCTGATGATTAC 357
Db 241 CGCATGGTATCCGCTCGGACGATTTGGGGCGGTACCGCTCTATCCCGGCTGATGATTAC 300
QY 358 GTCCCGCGGAATCCATTTGGCTGCTGCAGTACTCAATACCGGTTCATATTCGTGACAT 417
Db 301 GTCCCGCGGAATCCATTTGGCTGCTGCAGTACTCAATACCGGTTCATATTCGTGACAT 360
QY 418 CTGCGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGAC 477
Db 361 CTGCGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGAC 420
QY 478 CAAGAGCTGTGGTTCGCGCGGCGGCTCTTCGCGCGGAGCTTCATTCGCGCATCGTGCC 537
Db 421 CAAGAGCTGTGGTTCGCGCGGCGGCTCTTCGCGCGGAGCTTCATTCGCGCATCGTGCC 480
QY 538 TGGCTACATCAGTAGTTCGCTGGTGGATCGTACGATACGAGGCGCATTCGCATATTCGC 597
Db 481 TGGCTACATCAGTAGTTCGCTGGTGGATCGTACGATACGAGGCGCATTCGCATATTCGC 540
QY 598 CTTGCAGTTCACCTACTCTCTGCGGTGCGCTCAGTGAAGACTGGATCCGTTGTTCTGTC 657
Db 541 CTTGCAGTTCACCTACTCTCTGCGGTGCGCTCAGTGAAGACTGGATCCGTTGTTCTGTC 600
QY 658 GGCACACCGCGCTTT 672
Db 601 GGCACACCGCGCTTT 615
```

Search completed: December 16, 2004, 07:12:51

Job time : 11096.6 secs

Blank
page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 22:52:10 ; Search time 10054.9 Seconds
(without alignments)
11367.538 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctaagcgagagatgtctgc.....ccaaaaa..... 2417

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ats.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 2417 | 100.0 | 2417 | 3 | AF132552 |
| 2 | 2417 | 100.0 | 2417 | 6 | AX799088 |
| 3 | 2410.4 | 99.7 | 2699 | 6 | CQ589353 |
| C 4 | 2147.4 | 88.8 | 4922 | 6 | CQ589352 |
| C 5 | 2147.4 | 88.8 | 163921 | 3 | AC007853 |
| C 6 | 2147.4 | 88.8 | 181132 | 3 | AC008206 |
| C 7 | 2147.4 | 88.8 | 227219 | 3 | AE003750 |
| C 8 | 2131.6 | 88.2 | 75650 | 2 | AC018145 |
| 9 | 962.2 | 39.8 | 4236 | 10 | BC052433 |
| 10 | 960.6 | 39.7 | 2710 | 6 | AX799084 |
| 11 | 955.8 | 39.5 | 2481 | 6 | AX799082 |
| 12 | 955.8 | 39.5 | 2481 | 9 | AY074880 |
| 13 | 733.4 | 30.3 | 2955 | 6 | CQ573675 |
| 14 | 720.2 | 29.8 | 2608 | 5 | BC046072 |
| 15 | 718.6 | 29.7 | 2567 | 5 | BC063234 |
| C 16 | 715.2 | 29.6 | 110000 | 8 | Continuation (22 o |
| C 17 | 696 | 28.8 | 2953 | 6 | CQ596819 |
| 18 | 636 | 26.3 | 2118 | 9 | BT007100 |
| 19 | 636 | 26.3 | 2118 | 12 | BT008132 |

| | | | | | |
|------|-------|------|--------|----|-------------------|
| 20 | 636 | 26.3 | 2450 | 9 | BC048348 |
| 21 | 636 | 26.3 | 2458 | 6 | CQ834714 |
| 22 | 636 | 26.3 | 2472 | 6 | CQ727867 |
| 23 | 636 | 26.3 | 2516 | 9 | BC020965 |
| 24 | 636 | 26.3 | 2898 | 9 | BSM809254 |
| 25 | 632.8 | 26.2 | 2293 | 9 | HUMTMC |
| 26 | 632.8 | 26.2 | 2472 | 6 | AX799092 |
| 27 | 632.8 | 26.2 | 2472 | 9 | HUMBSA |
| 28 | 631.2 | 26.1 | 2736 | 10 | BC037612 |
| 29 | 631.2 | 26.1 | 3094 | 6 | AX799090 |
| 30 | 631.2 | 26.1 | 3094 | 10 | MJSTRPR |
| 31 | 628 | 26.0 | 2481 | 5 | BC067313 |
| 32 | 628 | 26.0 | 2481 | 10 | AK129027 |
| 33 | 623.2 | 25.9 | 2435 | 6 | CQ834710 |
| 34 | 615.6 | 25.5 | 2760 | 6 | CQ426624 |
| 35 | 615.6 | 25.5 | 2819 | 8 | AK067609 |
| 36 | 614 | 25.4 | 2756 | 8 | AK099138 |
| 37 | 607.4 | 25.1 | 1734 | 10 | BC003206 |
| 38 | 582.2 | 24.1 | 54118 | 3 | U13019 |
| 39 | 579.6 | 24.0 | 300029 | 8 | AE016903 |
| C 40 | 576.8 | 23.9 | 110000 | 8 | CR382123_09 |
| C 41 | 576.8 | 23.9 | 110000 | 8 | CR382123_10 |
| 42 | 567.4 | 23.5 | 2555 | 8 | BT012881 |
| 43 | 561.2 | 23.2 | 110000 | 8 | Continuation (1 o |
| 44 | 523.4 | 21.7 | 3633 | 8 | SCYGL022W |
| 45 | 521.8 | 21.6 | 2733 | 6 | AX799086 |

ALIGNMENTS

| | | | | | | |
|------------|---|-------------------------|-------------|-------|--------|-----------------|
| RESULT 1 | AF132552 | Drosophila melanogaster | 2417 bp | mRNA | linear | INV 12-FEB-2003 |
| AF132552 | Drosophila melanogaster | GM01838 | full insert | cDNA. | | |
| LOCUS | AF132552 | | | | | |
| DEFINITION | AF132552 | | | | | |
| ACCESSION | AF132552.1 | GI:4689327 | | | | |
| VERSION | FLI CDNA. | | | | | |
| KEYWORDS | Drosophila melanogaster (fruit fly) | | | | | |
| SOURCE | Drosophila melanogaster | | | | | |
| ORGANISM | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. | | | | | |
| REFERENCE | 1 (bases 1 to 2417) | | | | | |
| AUTHORS | Tsang G., Brokstein, P., Frise, E., Harvey, D., Evans-Holm, M., Lewis, S.E., Suh, C. and Rubin, G.M. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (02-MAR-1999) Berkeley Drosophila Genome Project, University of California Berkeley, 539 Life Sciences Addition #3200, Berkeley, CA 94720, USA | | | | | |
| REFERENCE | 2 (bases 1 to 2417) | | | | | |
| AUTHORS | Stapleton, M., Brokstein, P., Hong, L., Agbavani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Drensek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Kronmiller, B. Li, P., Liao, G., Miranda, A., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Phuanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Celniker, S. | | | | | |
| TITLE | Direct Submission | | | | | |
| JOURNAL | Submitted (12-FEB-2003) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA | | | | | |
| COMMENT | Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory Berkeley, CA 94720 | | | | | |

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal

priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES
SOURCE

RES
SOURCE

1 3417

/organism="Drosophila melanogaster"

```
/pol type="trna"
```

```
/db xref="taxon:7227"
```

<1.2417

```
/gene="OstStt3"
```

/db_xref="FLYBASE:FBgn0011336"

<1. .2395

```
/gene="OstStt3"
```

```
/note="Longest OKF"
```

"06E9LUMC-4"CM07838n/
xrxob11c1="CM07838n/
codon_bar=2

```

/produc= grove3ep
/protein id="ADP27851 2"

```

```
/proc/in_iq= iHEZ105I.  
/db xref="GI:28317200"
```

```
/db xref="FLYBASE:FBgn0011336"
```

/r=elation="LSBCEVFAQI GYNERFTWGNWDMRTPKMLASKVAGYSLLIT
 ALLIWLWAGSRLFAVIREZSI IHEFDPMFNEAYAYVWGVSGLTSTLYLLT
 PUGRIWVYFELMTSGGHLHLLWLVPHIRDICVFLAPISGLTSTLYLLT
 IESASAGAFNAASFTAVGYSIRSVAGSDNEGIALFAFOFTYFWKSVZGVFV
 SAAALXVYVWASVAGVYFINLIPHVLVWYSRPLATSYSTYVLCGLFSM
 QIPVGFQPIRTSEHAAVGFVLMVATVHULQUSVSESRKPLVIGVLLGVGV
 FVAVVYVTLWGVAPWSGRVFSLNDQYAKTHIIVASVEHQFTWFSFDDHLV
 CAPPVWYCIQINDERVVVLIVASVYAGVWYVRLMTLTPVVMCLAGVAFSGLL
 VLFQBSRGMGLTAACTEVEAEDSIEKTYDKAGKLKTHKDQDQDQGVSN
 LKSIIVALLMMLFPAHTEVNSYVSPVLAHPNSODGRLDQDQYVWL
 SONTADRAWNSWDYQIAGMARTPLDNNWNHSHLWVKAMSTREKSYIM
 TSLDVLVIFVGGVYSGDINKFLMVAIABEGHQIKESDYFTDGRFVDAE
 GAPALLCNMYLSYRGEKLDVSGSGDYRTNVAIKNQDFDLTLEBAVTHW
 VAPLVAIKVHEZBNPSIKWZBETPBNESVSPVSPVAPVAPVAPVAPVAPV

OBJET

| Query Match | 100.0%; Score 2417; DB 3; Length 2417; | |
|---|---|--|
| Best Local Similarity | 100.0%; Pred. No. 0; | |
| Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |
| QY | 1 TCTAAGCGAAGAAATGTGTCGTTGCAATTCAGATCGGTATAATATTCAGATTACTGGCTG 60 | |
| DB | 1 TCTAAGCGAAGAAATGTGTCGTTGCAATTCAGATCGGTATAATATTCAGATTACTGGCTG 60 | |
| QY | 61 GAATTGGGACATGAATCGGACCGCGAAGATCTGAACAGCAAGGTGGCTGCCTACAGCAG 120 | |
| DB | 61 GAATTGGGACATGAATCGGACCGCGAAGATCTGAAACAGCAAGGTGGCTGCCTACAGCAG 120 | |
| QY | 121 CCTAATCACCTTCCGCATCTCTGTGAATCGCTGGCTGGCGGGAATTTCTCTCGGCTCTT 180 | |
| DB | 121 CCTAATCACCTTCCGCATCTCTGTGAATCGCTGGCTGGCGGGAATTTCTCTCGGCTCTT 180 | |
| QY | 181 CGCGCTCATCGTTTCAGATCGAATATCCATGAGTTTGAATCCGTGGTTCAACTACCGGGC 240 | |
| DB | 181 CGCGCTCATCGTTTCAGATCGAATATCCATGAGTTTGAATCCGTGGTTCAACTACCGGGC 240 | |
| QY | 241 CACGCCCTACATGTCGAGAAATGGTTGGTACAACTTCCCTCAACTGTTTCGACGAGCGGC 300 | |
| DB | 241 CACGCCCTACATGTCGAGAAATGGTTGGTACAACTTCCCTCAACTGTTTCGACGAGCGGC 300 | |
| QY | 301 ATGGTATCCGCTCGGACAGGATGTGGGCGGTACCGCTATCCCGGCGCTGATGATTACGTC 360 | |
| DB | 301 ATGGTATCCGCTCGGACAGGATGTGGGCGGTACCGCTATCCCGGCGCTGATGATTACGTC 360 | |
| QY | 361 CGGCGGAATCAATTGGCTGCTGCAAGTACATACACCGGTCCTATTTGGTGAACATCTG 420 | |
| DB | 361 CGGCGGAATCAATTGGCTGCTGCAAGTACATACACCGGTCCTATTTGGTGAACATCTG 420 | |
| QY | 421 CGTGTCTCTCGGCGCGGATCTTCAGTGGCCTGAGCTCCATCTCCACCTACCTGCTGACCAA 480 | |

| | | | |
|------|----|---|------|
| 421 | DB | CGTGTTCTCGCGCGGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACCTGCTGACCAA | 480 |
| 481 | QY | GGAGCTGTGTGTCCGCGGCGCGGCGCTCTTCGCCGCCAGCTTTCATGCCATCGTGCCTGG | 540 |
| 481 | DB | GGAGCTGTGTGTCCGCGGCGCGGCGCTCTTCGCCGCCAGCTTTCATGCCATCGTGCCTGG | 540 |
| 541 | QY | CTACATCAGTAGGTGGTGGCTGGATCGTAGATAAAGAGGGCATTCGCATATCGCCCT | 600 |
| 541 | DB | CTACATCAGTAGGTGGTGGCTGGATCGTAGATAAAGAGGGCATTCGCATATCGCCCT | 600 |
| 601 | QY | GCAGTTCACCTACTTCTCTGTGGTGGCTCAGTAGAAGACTGGATCCGTGTTCGTGCGC | 660 |
| 601 | DB | GCAGTTCACCTACTTCTCTGTGGTGGCTCAGTAGAAGACTGGATCCGTGTTCGTGCGC | 660 |
| 661 | QY | CGCAGCGCTTGTTCCTACTCTTCAATGCTGCCGTCCGCTGGGTGGTACGTGTTCATCAT | 720 |
| 661 | DB | CGCAGCGCTTGTTCCTACTCTTCAATGCTGCCGTCCGCTGGGTGGTACGTGTTCATCAT | 720 |
| 721 | QY | CAACCTGATACCCCTGACGCTTCCTGATCTCTCATTTATGGGAGGTACTCGCCGCGTCT | 780 |
| 721 | DB | CAACCTGATACCCCTGACGCTTCCTGATCTCTCATTTATGGGAGGTACTCGCCGCGTCT | 780 |
| 781 | QY | GCTGACCAAGCTACAGCACCTTCTACATCTCGGACTGCTGTTCTTCATGACAGATCCCTT | 840 |
| 781 | DB | GCTGACCAAGCTACAGCACCTTCTACATCTCGGACTGCTGTTCTTCATGACAGATCCCTT | 840 |
| 841 | QY | CGTGGGATTCGAACCGATACGCACAGTGAACAATGGCTCGCTGGGAGTGTGTGTCT | 900 |
| 841 | DB | CGTGGGATTCGAACCGATACGCACAGTGAACAATGGCTCGCTGGGAGTGTGTGTCT | 900 |
| 901 | QY | CCTTATGGCGGTGGCCACCTTCGCGCAATTCGACGTCCGTGCTGCGGCAACAGAGTCCG | 960 |
| 901 | DB | CCTTATGGCGGTGGCCACCTTCGCGCAATTCGACGTCCGTGCTGCGGCAACAGAGTCCG | 960 |
| 961 | QY | GAAGCTGTTCATGTCGCGCGGATGCTGGTGGGCGTTGGCGTTTGGCGTCTGTGGCGCTCGTGT | 1020 |
| 961 | DB | GAAGCTGTTCATGTCGCGCGGATGCTGGTGGGCGTTGGCGTTTGGCGTCTGTGGCGCTCGTGT | 1020 |
| 1021 | QY | GCTCACCATGCTGGGCGTTGTGGCCCGGTGAGTGGACGCTTCTACTCGCTGTGGGATAC | 1080 |
| 1021 | DB | GCTCACCATGCTGGGCGTTGTGGCCCGGTGAGTGGACGCTTCTACTCGCTGTGGGATAC | 1080 |
| 1081 | QY | TGGCTACGCCAAGATCCAGATTCGCATCATTCGATTCGGTTCGGAGGATCAGCCACACAC | 1140 |
| 1081 | DB | TGGCTACGCCAAGATCCAGATTCGCATCATTCGATTCGGTTCGGAGGATCAGCCACACAC | 1140 |
| 1141 | QY | TTGGTTCTCGTTCTTCTTTGATCTGCACATCCTGGTGTGCGCTTCCCACTGGAGTGTG | 1200 |
| 1141 | DB | TTGGTTCTCGTTCTTCTTTGATCTGCACATCCTGGTGTGCGCTTCCCACTGGAGTGTG | 1200 |
| 1201 | QY | GTACTGATCAAGAGATCAACGACGAGCGGTTTTCTGGTGGCTGTACGCCATCAGTGC | 1260 |
| 1201 | DB | GTACTGATCAAGAGATCAACGACGAGCGGTTTTCTGGTGGCTGTACGCCATCAGTGC | 1260 |
| 1261 | QY | GGTTTACTTCGCTGGTGTGATGGTGGCTTTCATGCTGACCTCAGCCGGTGGTGTGCAT | 1320 |
| 1261 | DB | GGTTTACTTCGCTGGTGTGATGGTGGCTTTCATGCTGACCTCAGCCGGTGGTGTGCAT | 1320 |
| 1321 | QY | GCTGGCCGAGTGGCTTTTTCGGGACTGTTGGATGTGTTCTCTGGAAGAGGATTCGTCCTAA | 1380 |
| 1321 | DB | GCTGGCCGAGTGGCTTTTTCGGGACTGTTGGATGTGTTCTCTGGAAGAGGATTCGTCCTAA | 1380 |
| 1381 | QY | GCGAATGGGCACAGCCATTAAGCGCACCCGAAGTGGATGAAGCTGAGGATTCATTGA | 1440 |
| 1381 | DB | GCGAATGGGCACAGCCATTAAGCGCACCCGAAGTGGATGAAGCTGAGGATTCATTGA | 1440 |
| 1441 | QY | GAAGAAGACGCTGTACGACAGCGTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCA | 1500 |
| 1441 | DB | GAAGAAGACGCTGTGTACGACAGCGTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCA | 1500 |
| 1501 | QY | GCAGGATCTGGCGTTCAGCTCCAACTGAAGAGTATTTGTTATTTTGGCGGTTCTAATGCT | 1560 |
| 1501 | DB | GCAGGATCTGGCGTTCAGCTCCAACTGAAGAGTATTTGTTATTTTGGCGGTTCTAATGCT | 1560 |

QY 1561 GTTGATGATGTTGGTGTCCACTGCACTGGGTGACAGCAATGCCCTACTCCAGTCCCTC 1620
 Db 1561 GTTGATGATGTTGGTGTCCACTGCACTGGGTGACAGCAATGCCCTACTCCAGTCCCTC 1620
 QY 1621 CATTGCTGTTGTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTACAGCATTTTCAG 1680
 Db 1621 CATTGCTGTTGTTTCCCAACAGTCAAGATGGATCCCGCAACATTTTACAGCATTTTCAG 1680
 QY 1681 AGAGGCTTACTACTGCTTTCCAGAAACACTGCGCATGATGCTCGGTTATGCTTGGTG 1740
 Db 1681 AGAGGCTTACTACTGCTTTCCAGAAACACTGCGCATGATGCTCGGTTATGCTTGGTG 1740
 QY 1741 CGATTACGGATACACAGATAGCGGAAATGCGAAACAGAACGACGCTAGTGATATATATAC 1800
 Db 1741 CGATTACGGATACACAGATAGCGGAAATGCGAAACAGAACGACGCTAGTGATATATATAC 1800
 QY 1801 GTGGAAACATAGTCAATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAAATC 1860
 Db 1801 GTGGAAACATAGTCAATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAAATC 1860
 QY 1861 CTACGAAATATGACATCTCTGACCTGGACTAGCTTTTGGTGATCTTTGGCGGTGTAT 1920
 Db 1861 CTACGAAATATGACATCTCTGACCTGGACTAGCTTTTGGTGATCTTTGGCGGTGTAT 1920
 QY 1921 CGGCTATTCTGCGATGATATCAACAGTTCCTGTGGATGGTCCGAATTTGCTGAGGAGA 1980
 Db 1921 CGGCTATTCTGCGATGATATCAACAGTTCCTGTGGATGGTCCGAATTTGCTGAGGAGA 1980
 QY 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTACCGACCGCGGTGAAATTCAGGGTGA 2040
 Db 1981 GCATCCCAAGGACATTAAGGAAAGGATTAATTTACCGACCGCGGTGAAATTCAGGGTGA 2040
 QY 2041 TGCCGAAGGTGCTCGGCGCTGCTCACTGCGCTTATGTACAAATTAAGCTACTACAGATT 2100
 Db 2041 TGCCGAAGGTGCTCGGCGCTGCTCACTGCGCTTATGTACAAATTAAGCTACTACAGATT 2100
 QY 2101 CGCGGAATTTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACACGTAACGCGT 2160
 Db 2101 CGCGGAATTTGAAGTTGGACTACAGAGTCCATCTGGATATGATCGCACACGTAACGCGT 2160
 QY 2161 CATTGGGAATAAGACTTGGATCTGACCTACCTGAGAGAGGCTTACACCAAGCAACTG 2220
 Db 2161 CATTGGGAATAAGACTTGGATCTGACCTACCTGAGAGAGGCTTACACCAAGCAACTG 2220
 QY 2221 GCTTGTTCGATCTATAGGTGAAGAGCCGATGATTCATATAGACCATCCTGAGAGAC 2280
 Db 2221 GCTTGTTCGATCTATAGGTGAAGAGCCGATGATTCATATAGACCATCCTGAGAGAC 2280
 QY 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTGGCAA 2340
 Db 2281 CAAGGAGAGACGATTCCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTGGCAA 2340
 QY 2341 GGGCTACATACGAAACCGACCGGTTGTTTAAAGGAAACGAAACCTTGAATAAACCAC 2400
 Db 2341 GGGCTACATACGAAACCGACCGGTTGTTTAAAGGAAACGAAACCTTGAATAAACCAC 2400
 QY 2401 AAAAAAATAAAAAA 2417
 Db 2401 AAAAAAATAAAAAA 2417

RESULT 2
 AX799088 2417 bp DNA linear PAT 08-OCT-2003
 LOCUS Sequence 7 from Patent WO03054008.
 DEFINITION AX799088
 ACCESSION AX799088
 VERSION AX799088.1 GI:37605061
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.
 1
 Perreault, C. and McBride, K.
 Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy
 Patent: WO 03054008-A 7 03-JUL-2003;
 Compatigene Inc. (CA)
 Location/Qualifiers
 1. 2417
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /note="STT3, Acc# AF132552"
 ORIGIN
 Query Match 100.0%; Score 2417; DB 6; Length 2417;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCTAAGCGAAGAATGTGTGCTTGCATTTTCAGATCGGTTAATAATTTTCGAGTTACTGCGTG 60
 Db 1 TCTAAGCGAAGAATGTGTGCTTGCATTTTCAGATCGGTTAATAATTTTCGAGTTACTGCGTG 60
 QY 61 GAATTTGGACATGAATCGGACCGGACAGATGCTGACAGCAAGTGGCTGCTACAGCAG 120
 Db 61 GAATTTGGACATGAATCGGACCGGACAGATGCTGACAGCAAGTGGCTGCTACAGCAG 120
 QY 121 CCTAATACCTTCGCCATCTCTGCTAATCGCTGGCTGGCGGATTTTCTCTCGCCTCTT 180
 Db 121 CCTAATACCTTCGCCATCTCTGCTAATCGCTGGCTGGCGGATTTTCTCTCGCCTCTT 180
 QY 181 CGCGTCACTCGTTTCGATGATTCATGATTTGATTCGATTCGATTCGATTCGATTCGATTC 240
 Db 181 CGCGTCACTCGTTTCGATGATTCATGATTTGATTCGATTCGATTCGATTCGATTCGATTC 240
 QY 241 CACGCGCTACATGCTGCGAGATGCTTGTACAACTTCTCAACTGTTTTCGACAGGCGCG 300
 Db 241 CACGCGCTACATGCTGCGAGATGCTTGTACAACTTCTCAACTGTTTTCGACAGGCGCG 300
 QY 301 ATGCTATCCGCTCGGACGATTTGGCGGTACCGTCTATCCGCGCTGATGATGATGATG 360
 Db 301 ATGCTATCCGCTCGGACGATTTGGCGGTACCGTCTATCCGCGCTGATGATGATGATGATG 360
 QY 361 CGGCGGAATCATTGGTGTGCTGCACTGCTGCTCAACATACCGTTCATATTCGTCGATCTG 420
 Db 361 CGGCGGAATCATTGGTGTGCTGCTGCTCAACATACCGTTCATATTCGTCGATCTG 420
 QY 421 CGTGTTCCTGCGCGGATCTTCAGTGGCTGACCTGCTCACTCCACTGCTGCTGACCA 480
 Db 421 CGTGTTCCTGCGCGGATCTTCAGTGGCTGACCTGCTCACTCCACTGCTGCTGACCA 480
 QY 481 GGAGCTGTGCTCGCGCGGCGCGGCTCTTTCGCGCGGCGGCTTCATCGCCATCGTGCCTG 540
 Db 481 GGAGCTGTGCTCGCGCGGCGCGGCTCTTTCGCGCGGCGGCTTCATCGCCATCGTGCCTG 540
 QY 541 CTACATCAGTAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 Db 541 CTACATCAGTAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 QY 601 GCAGTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 Db 601 GCAGTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 QY 661 CGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 Db 661 CGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 QY 721 CAACCTGATACCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 Db 721 CAACCTGATACCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
 QY 781 GCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

Db 781 GCTGACACGCTACAGCACCTTCTACATCCTGGGACTGCTGTTCTCCATCAGATCCCTTT 840
 Qy 841 CGTGGGATCCAAACCGATACGACCAAGTGAACATGCTGCTGGGAGTGTGTGCT 900
 Db 841 CGTGGGATCCAAACCGATACGACCAAGTGAACATGCTGGGAGTGTGTGCT 900
 Qy 901 CTTTATGGCCGCTGGCCACCTTGGCCATTTGCAAGTCCGCTGCTGCGCAAGAGTTCGG 960
 Db 901 CTTTATGGCCGCTGGCCACCTTGGCCATTTGCAAGTCCGCTGCTGCGCAAGAGTTCGG 960
 Qy 961 GAAGCTGTTCTATCTGCGCGGATTTGCTGTTGGGCGTGTGGGCTTTGCTGCGCTGCTGCT 1020
 Db 961 GAAGCTGTTCTATCTGCGCGGATTTGCTGTTGGGCGTGTGGGCTTTGCTGCGCTGCTGCT 1020
 Qy 1021 GCTCACCATGCTGGGCGTGTGGGCGGCTGGGAGTGGAGCGCTTCTACTCGCTGCTGGGATAC 1080
 Db 1021 GCTCACCATGCTGGGCGTGTGGGCGGCTGGGAGTGGAGCGCTTCTACTCGCTGCTGGGATAC 1080
 Qy 1081 TGGCTACGCCAAGATCCACATCCCATCATTTGCTGCTGGGAGCATCAGCCACAC 1140
 Db 1081 TGGCTACGCCAAGATCCACATCCCATCATTTGCTGCTGGGAGCATCAGCCACAC 1140
 Qy 1141 TTGGTCTCTGCTTCTTCTTCTGATCTGCAATCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 Db 1141 TTGGTCTCTGCTTCTTCTTCTGATCTGCAATCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 Qy 1201 GTACTGCAATCAAGCAGATCAACGACGAGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 1201 GTACTGCAATCAAGCAGATCAACGACGAGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Qy 1261 GCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Db 1261 GCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Qy 1321 GCTGCGGAGTGGCTTTTGGGACTGTTGGAGTGTTCCTGCAAGAGATTCGCTAA 1380
 Db 1321 GCTGCGGAGTGGCTTTTGGGACTGTTGGAGTGTTCCTGCAAGAGATTCGCTAA 1380
 Qy 1381 GCGAATGGGACAGCCATTAAGCGCAGCCAGCGAGTGGATGAAGCTGAGGATTCATTTGA 1440
 Db 1381 GCGAATGGGACAGCCATTAAGCGCAGCCAGCGAGTGGATGAAGCTGAGGATTCATTTGA 1440
 Qy 1441 GAAGAGAGCTGTACGACAGGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCA 1500
 Db 1441 GAAGAGAGCTGTACGACAGGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCA 1500
 Qy 1501 GCAGGATACCTGGCTGAGCTCCACCTGCAAGAGTATTTGTTATTTGGCGCTTCTAATGCT 1560
 Db 1501 GCAGGATACCTGGCTGAGCTCCACCTGCAAGAGTATTTGTTATTTGGCGCTTCTAATGCT 1560
 Qy 1561 GTTGTATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 1561 GTTGTATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Qy 1621 CATTTCTTTGCTTTTCCACACAGTCAAGATGATGATGATGATGATGATGATGATGATGATG 1680
 Db 1621 CATTTCTTTGCTTTTCCACACAGTCAAGATGATGATGATGATGATGATGATGATGATGATG 1680
 Qy 1681 AGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
 Db 1681 AGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
 Qy 1741 GGATTACGGATACCGGATGCGGATGCGCAACAGACGAGTGTGTGTGTGTGTGTGTGTGTGT 1800
 Db 1741 GGATTACGGATACCGGATGCGGATGCGCAACAGACGAGTGTGTGTGTGTGTGTGTGTGTGT 1800
 Qy 1801 GTGGAACAATAGTCAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 Db 1801 GTGGAACAATAGTCAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 Qy 1861 CTACGAATATGATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
 Db 1861 CTACGAATATGATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920

Qy 1921 CGGCTATTCTGGGATGATATCAACAAGTCTCTGCTGGATGCTCGAATTCGTGAGGAGA 1980
 Db 1921 CGGCTATTCTGGGATGATATCAACAAGTCTCTGCTGGATGCTCGAATTCGTGAGGAGA 1980
 Qy 1981 GCATCCCAAGGACATTAAGGAAAGCATTTACTTTTACCGACCGCGTGAATTCAGGGTGA 2040
 Db 1981 GCATCCCAAGGACATTAAGGAAAGCATTTACTTTTACCGACCGCGTGAATTCAGGGTGA 2040
 Qy 2041 TGCCGAAGTGTCTCGGCGCTGCTCAATGCTGCTTATGTACAAATTAAGCTACTACAGATT 2100
 Db 2041 TGCCGAAGTGTCTCGGCGCTGCTCAATGCTGCTTATGTACAAATTAAGCTACTACAGATT 2100
 Qy 2101 CGGGGAATTTGAAGTTGGACTACAGAGTCCATCTGATATGATGATGATGATGATGATGATG 2160
 Db 2101 CGGGGAATTTGAAGTTGGACTACAGAGTCCATCTGATATGATGATGATGATGATGATGATG 2160
 Qy 2161 CATTTGGGATTAAGGATTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Db 2161 CATTTGGGATTAAGGATTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Qy 2221 GCTTGTTCGATCTATAGGGTGAAGACCGCATGAGTTCAATAGACCATCACTGAAGAC 2280
 Db 2221 GCTTGTTCGATCTATAGGGTGAAGACCGCATGAGTTCAATAGACCATCACTGAAGAC 2280
 Qy 2281 CAAGGAGACAGATTCCTCCAGCAAACTTCATTTCCGAGGAAGAACTCTAAGCGTCCGAA 2340
 Db 2281 CAAGGAGACAGATTCCTCCAGCAAACTTCATTTCCGAGGAAGAACTCTAAGCGTCCGAA 2340
 Qy 2341 GGCTACATACGAAACCGACCGCTTGTGTTAAGGAAAAAGAACTTGAATAAATTAACCCCA 2400
 Db 2341 GGCTACATACGAAACCGACCGCTTGTGTTAAGGAAAAAGAACTTGAATAAATTAACCCCA 2400
 Qy 2401 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2417
 Db 2401 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2417

RESULT 3
 LOCUS CQ589353 2699 bp DNA linear PAT 02-FEB-2004
 DEFINITION Sequence 17111 from Patent WO0171042.
 ACCESSION CQ589353
 VERSION CQ589353.1 GI:41648215
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1
 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 TITLE Detection kit, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 JOURNAL Patent: WO 0171042-A 17111 27-SEP-2001;
 PE Corporation (NY) (US)

FEATURES
 Location/Qualifiers
 source 1..2699
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"

ORIGIN
 Query Match 99.7%; Score 2410.4; DB 6; Length 2699;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2411; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCTAACCGAAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
 Db 10 TCTAACCGAAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 69
 Qy 61 GAATTGGGACATGAATCGGACCGCAAGATGCTGAACGAGAGTGGCTGGCTACAGCAG 120

Db 70 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAAACAGCAAGGTGGCTGCTACAGCAG 129
Qy 121 CTAATATCACTTCGCGCATCTCTGTAATCGCTTGGCTGGCGGATTTTCTCTCTCGCCTCTT 180
Db 130 CTAATATCACTTCGCGCATCTCTGTAATCGCTTGGCTGGCGGATTTTCTCTCTCGCCTCTT 189
Qy 181 CCGCTCATCGCTTCGAGTCGATTCATCAATGATTTGATCGGTGCTCAACTACCGGCG 240
Db 190 CGCGCTCATCGCTTCGAGTCGATTCATCAATGATTTGATCGGTGCTCAACTACCGGCG 249
Qy 241 CACCGCTCATCGCTTCGAGTCGATTTGATGTAACATTCCTCACTGGCTCGACGAGCGCG 300
Db 250 CACCGCTCATCGCTTCGAGTCGATTTGATGTAACATTCCTCACTGGCTCGACGAGCGCG 309
Qy 301 ATGGTATCGCTTCGCGCATTTGGCGGATTCGCGGTACCGTTCATCCCGCGCTGATGATTCGTC 360
Db 310 ATGGTATCGCTTCGCGCATTTGGCGGATTCGCGGTACCGTTCATCCCGCGCTGATGATTCGTC 369
Qy 361 CGCGGAAATCCATTTGGCTTCGACGTAATCAACATACCGGTCCATATTCGTGACATCTG 420
Db 370 CGCGGAAATCCATTTGGCTTCGACGTAATCAACATACCGGTCCATATTCGTGACATCTG 429
Qy 421 CGTGTCTCGCGCGGATTCCTGAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 480
Db 430 CGTGTCTCGCGCGGATTCCTGAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAA 489
Qy 481 GGAGCTGTGGTCCGCGCGCGCTCTTCGCGCGGATTCATTCGCGCATCTGCTGG 540
Db 490 GGAGCTGTGGTCCGCGCGGCTCTTCGCGCGGATTCATTCGCGCATCTGCTGG 549
Qy 541 CTACATCAGTAGTTCGCTGGATCGTACGATTAACAGGGCAATTCGCCATATTCGCCCT 600
Db 550 CTACATCAGTAGTTCGCTGGATCGTACGATTAACAGGGCAATTCGCCATATTCGCCCT 609
Qy 601 GCAGTTCACCTACTCTGCTGGTGGCTCAGTGAAGACTGGATCGGTTCCTGCTGGCG 660
Db 610 GCAGTTCACCTACTCTGCTGGTGGCTCAGTGAAGACTGGATCGGTTCCTGCTGGCG 669
Qy 661 CGCAGCGCTTTGCTACTTCTACATGGTGTTCGCGCTGGGTGGCTTACGTTTCATCAT 720
Db 670 CGCAGCGCTTTGCTACTTCTACATGGTGTTCGCGCTGGGTGGCTTACGTTTCATCAT 729
Qy 721 CACCTCATACCTTCGACCTTCGCTACCTGCTGCTATTCGCGGAGTACCTGCGCGCT 780
Db 730 CACCTCATACCTTCGACCTTCGCTACCTGCTGCTATTCGCGGAGTACCTGCGCGCT 789
Qy 781 GCTGACGAGTACAGCACTTCTACATCTCGGACTGCTGCTTCTCCATCGACATCCCTT 840
Db 790 GCTGACGAGTACAGCACTTCTACATCTCGGACTGCTGCTTCTCCATCGACATCCCTT 849
Qy 841 CGTGGGATTCACCGGATACGACCGAGTGAACATGCTGCTGGCTGGAGTGTGTGCT 900
Db 850 CGTGGGATTCACCGGATACGACCGAGTGAACATGCTGCTGGCTGGAGTGTGTGCT 909
Qy 901 CCTTATGGCGCTGGCGCACTTCGCGCATTTGCACTCGCTGCTGCTGCGCAACGAGTTCG 960
Db 910 CCTTATGGCGCTGGCGCACTTCGCGCATTTGCACTCGCTGCTGCTGCGCAACGAGTTCG 969
Qy 961 GAAGCTGTTTCATCGTTCGCGGATTTGCTGGCGCTTGGCGCTTGGCGCTGCTGCT 1020
Db 970 GAAGCTGTTTCATCGTTCGCGGATTTGCTGGCGCTTGGCGCTTGGCGCTGCTGCT 1029
Qy 1021 GCTCACCATGCTGGCGGTTGGCGCGGATTTGCTGGCGCTTGGCGCTTGGCGCTGCTGCT 1080
Db 1030 GCTCACCATGCTGGCGGTTGGCGCGGATTTGCTGGCGCTTGGCGCTTGGCGCTGCTGCT 1089
Qy 1081 TGCTACGCGCAAGATCCACATTCACATTCGATTCGCTGCTGGAGCATCAGCCACCCAC 1140
Db 1090 TGCTACGCGCAAGATCCACATTCACATTCGATTCGCTGCTGGAGCATCAGCCACCCAC 1149
Qy 1141 TTGGTTCGCTTCTTTGATTCGACATCTGCTGCTGGCGCTTCCAGTGGAGTGTG 1200
Db 1150 TTGGTTCGCTTCTTTGATTCGACATCTGCTGCTGGCGCTTCCAGTGGAGTGTG 1209

Qy 1201 GTACTGCATCAACAGCAGATCAACGACGAGCGCTTTTCGTTGGTCTGTACGCCATCAGTGC 1260
Db 1210 GTACTGCATCAACAGCAGATCAACGACGAGCGCTTTTCGTTGGTCTGTACGCCATCAGTGC 1269
Qy 1261 GGTTTACTTTCGCTGGTGTGATGGTGGTGTGATGTTGATGTTGACCCCTCACGCCGCTGGTGTGCAT 1320
Db 1270 GGTTTACTTTCGCTGGTGTGATGGTGGTGTGATGTTGATGTTGACCCCTCACGCCGCTGGTGTGCAT 1329
Qy 1321 GCTGGCGGAGTGGCGCTTTTCGGGACTGTTGGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 1380
Db 1330 GCTGGCGGAGTGGCGCTTTTCGGGACTGTTGGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 1389
Qy 1381 GCGAATGGGACAGCCATTAAGCGGACGACCAAGGATGATGAAGCTGAGGATTCATTTGA 1440
Db 1390 GCGAATGGGACAGCCATTAAGCGGACGACCAAGGATGATGAAGCTGAGGATTCATTTGA 1449
Qy 1441 GAAGAGAGCTGTGACGACCAAGCTGGGAGCTGAGCATGCTAATAGCATGATGCCCA 1500
Db 1450 GAAGAGAGCTGTGACGACCAAGCTGGGAGCTGAGCATGCTAATAGCATGATGCCCA 1509
Qy 1501 GCAGGATATCGCGCTCAGCTCCAACTGAAAGATGTTGTTATTTTGGCGCTTCTAATGCT 1560
Db 1510 GCAGGATATCGCGCTCAGCTCCAACTGAAAGATGTTGTTATTTTGGCGCTTCTAATGCT 1569
Qy 1561 GTTGATGATTTTCGCTGCTCCACTGCACTGGGTGACCAAGCATGCTTACTCCAGTCCCTC 1620
Db 1570 GTTGATGATTTTCGCTGCTCCACTGCACTGGGTGACCAAGCATGCTTACTCCAGTCCCTC 1629
Qy 1621 CATTGCTTGGCTTTCCAAACAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Db 1630 CATTGCTTGGCTTTCCAAACAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1689
Qy 1681 AGAGCTTACTACTGCTTTCCGAGAACACTGCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1690 AGAGCTTACTACTGCTTTCCGAGAACACTGCGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1749
Qy 1741 GGAATACCGATACAGATAGCGGATGCGGAAACAGAACAGACGAGCTAGTGGATTAATATAC 1800
Db 1750 GGAATACCGATACAGATAGCGGATGCGGAAACAGAACAGACGAGCTAGTGGATTAATATAC 1809
Qy 1801 GTGGAAATATAGTACATAGCGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAGTGC 1860
Db 1810 GTGGAAATATAGTACATAGCGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAGTGC 1869
Qy 1861 CTACGAAATATAGTACATCTTTCGCTGGAGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1920
Db 1870 CTACGAAATATAGTACATCTTTCGCTGGAGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1929
Qy 1921 CGGCTATTTCGGCGATGATATCAACAGTTCCTGCTGGATGCTGCGAATGCTGAGGAGGAGA 1980
Db 1930 CGGCTATTTCGGCGATGATATCAACAGTTCCTGCTGGATGCTGCGAATGCTGAGGAGGAGA 1989
Qy 1981 GCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGA 2040
Db 1990 GCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGA 2049
Qy 2041 TGCGAAGGTGCTCCGCGCTGCTCACTGCTGCTTATGTAAGATTAAGCTACTACAGATT 2100
Db 2050 TGCGAAGGTGCTCCGCGCTGCTCACTGCTGCTTATGTAAGATTAAGCTACTACAGATT 2109
Qy 2101 CGGGAATTAAGCTTCCGATCTACAGAGTCCATCTGGATATGATGCTGACACAGTAAACGCGCT 2160
Db 2110 CGGGAATTAAGCTTCCGATCTACAGAGTCCATCTGGATATGATGCTGACACAGTAAACGCGCT 2169
Qy 2161 CATTGGGAATTAAGCTTCCGATCTACAGAGTCCATCTGGATATGATGCTGACACAGTAAACGCGCT 2220
Db 2170 CATTGGGAATTAAGCTTCCGATCTACAGAGTCCATCTGGATATGATGCTGACACAGTAAACGCGCT 2229
Qy 2221 GCTTGTTCGCTTCTATAGGCTGAAGAGCGCGATGATGCTCAATAGACCTACCTGAGAC 2280
Db 2230 GCTTGTTCGCTTCTATAGGCTGAAGAGCGCGATGATGCTCAATAGACCTACCTGAGAC 2289

QY 2281 CAAGGAGAGACGATTCTCTCAGCAAACTTTCATTTTCGAGAAAGAACTTAAGCGTCGCAA 2340
 Db 2290 CAAGGAGAGACGATTCTCTCAGCAAACTTTCATTTTCGAGAAAGAACTTAAGCGTCGCAA 2349
 QY 2341 GGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAAGAAAGAACTTGAATAAAGCCAA 2400
 Db 2350 GGGCTACATACGAAACCGACCGGTTGTTGTTAAGGAAAGAAAGAACTTGAATAAAGCCAA 2409
 QY 2401 AAAAAAATAA 2412
 Db 2410 AAAAAAATAA 2421
 RESULT 4
 CQ589352/c 4922 bp DNA linear PAT 02-FEB-2004
 LOCUS Sequence 17110 from Patent WO0171042.
 DEFINITION CQ589352
 ACCESSION CQ589352
 VERSION CQ589352.1 GI:41648214
 KEYWORDS
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1
 Venter, J.C.; Adams, M., Li, P.W. and Myers, E.W.
 Detection kits, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 Patent: WO 0171042-A 17110 27-SEP-2001;
 JOURNAL PE Corporation (NY) (US)
 FEATURES
 source Location/Qualifiers
 1..4922
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"
 ORIGIN
 Query Match 88.8%; Score 2147.4; DB 6; Length 4922;
 Best Local Similarity 91.5%; Pred. No. 0;
 Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;
 QY 1 TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 60
 Db 3913 TCTAAGCGAAGATGTCGTTGCAATTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 3854
 QY 61 GAATTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGGTGGCTGCTACAGCAG 120
 Db 3853 GAATTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGGTGGCTGCTACAGCAG 3794
 QY 121 CCTAATCACCTTCGCACTCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 180
 Db 3793 CCTAATCACCTTCGCACTCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 3734
 QY 181 CGCGCTCATCCGTTTCGAGTCGATTAATCCATGAGTTTGAATCCGTTCACTACCGGCG 240
 Db 3733 CGCGCTCATCCGTTTCGAGTCGATTAATCCATGAGTTTGAATCCGTTCACTACCGGCG 3674
 QY 241 CACGCGCTACATGGTCGAGATGGTTGGTACACTTCTCACTGGTTTCGACGCGCGC 300
 Db 3673 CACGCGCTACATGGTCGAGATGGTTGGTACACTTCTCACTGGTTTCGACGCGCGC 3614
 QY 301 ATGTATCCGCTCGGACGAGTTGGGCGGTACCGCTCTATCCCGGCTCATGATACGTC 360
 Db 3613 ATGTATCCGCTCGGACGAGTTGGGCGGTACCGCTCTATCCCGGCTCATGATACGTC 3554
 QY 361 CCGCGGAATCATGGCTGCTGACGCTACATACCGGTCCATATTCGTGACATCTG 420
 Db 3553 CCGCGGAATCATGGCTGCTGACGCTACATACCGGTCCATATTCGTGACATCTG 3494
 QY 421 CGTGTCTCTGGCGCGGATCTTTCAGTGGCCTGACCTCCATCTCCACCTACTCTGTGACCAA 480
 Db 3493 CGTGTCTCTGGCGCGGATCTTTCAGTGGCCTGACCTCCATCTCCACCTACTCTGTGACCAA 3434

QY 481 GGAGCTGTGTCCGCGGCGCGGCTCTTTCGCGCGCAGCTTTCATTCGCGCATCTGTCCTGG 540
 Db 3433 GGAGCTGTGTCCGCGGCGCGGCTCTTTCGCGCGCAGCTTTCATTCGCGCATCTGTCCTGG 3374
 QY 541 CTACATCAGTAGTTCGCTGGCTGGATCGTACGATAACGAGGCGCATTTGCCATATTTCGCGCT 600
 Db 3373 CTACATCAGTAGTTCGCTGGCTGGATCGTACGATAACGAGGCGCATTTGCCATATTTCGCGCT 3314
 QY 601 GCAGTTCACTACTTCTCTGTGGGTGCTCAGTGAAGACTGGAATCCGTTGTTCTGGTGGC 660
 Db 3313 GCAGTTCACTACTTCTCTGTGGGTGCTCAGTGAAGACTGGAATCCGTTGTTCTGGTGGC 3254
 QY 661 CGAGCGGCTTTGCTCTACTTCTACATGGTGCCTCCGCTGGGTGGCTACGTTGTTTCATCAT 720
 Db 3253 CGAGCGGCTTTGCTCTACTTCTACATGGTGCCTCCGCTGGGTGGCTACGTTGTTTCATCAT 3194
 QY 721 CAACCTGATACCCCTGCAGCTCTTCGTACTGCTCATATTATGGCAGGTACTCGCGCGTCT 780
 Db 3193 CAACCTGATACCCCTGCAGCTCTTCGTACTGCTCATATTATGGCAGGTACTCGCGCGTCT 3134
 QY 781 GCTGACCACTACAGCACTTCTACATCTGGGAGTCTGTTCTCCATGCGAGATCCCTT 840
 Db 3133 GCTGACCACTACAGCACTTCTACATCTGGGAGTCTGTTCTCCATGCGAGATCCCTT 3074
 QY 841 CGTGGGATTCCAAACCGATACGCAACAGTGAACACATGCTGCGCTGGGAGTGTGTTGCT 900
 Db 3073 CGTGGGATTCCAAACCGATACGCAACAGTGAACACATGCTGCGCTGGGAGTGTGTTGCT 3014
 QY 901 CCTTATGGCGGTGGCCACTTTCGCGCATTTGAGTCCGCTGCTGTCGCGCAAGAGTTCGG 960
 Db 3013 CCTTATGGCGGTGGCCACTTTCGCGCATTTGAGTCCGCTGCTGTCGCGCAAGAGTTCGG 2954
 QY 961 GAAGCTGTTTCATGCTCGGCGGATTTGCTGCTGGCGGTTCGCTGCTGCTGCTGCTGCT 1020
 Db 2953 GAAGCTGTTTCATGCTCGGCGGATTTGCTGCTGGCGGTTCGCTGCTGCTGCTGCTGCT 2894
 QY 1021 GCTCACCATGCTGGGCGTTTGTGCGCCGCTGGAGTGAAGCTTCTACTCTGCTGGGATAC 1080
 Db 2893 GCTCACCATGCTGGGCGTTTGTGCGCCGCTGGAGTGAAGCTTCTACTCTGCTGGGATAC 2834
 QY 1081 TGCTACGCCAAGATCCATTTCCCATTCATTCGATCCGCTGGAGCATACGCCACAC 1140
 Db 2833 TGCTACGCCAAGATCCATTTCCCATTCATTCGATCCGCTGGAGCATACGCCACAC 2774
 QY 1141 TTGCTTCTCGTTCTTTTGTGATCTGCACATCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 Db 2773 TTGCTTCTCGTTCTTTTGTGATCTGCACATCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2714
 QY 1201 GTACTGATCAAGAGATCAACGACGCGCTTTTCGTTGGTCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 2713 GTACTGATCAAGAGATCAACGACGCGCTTTTCGTTGGTCTGCTGCTGCTGCTGCTGCTGCT 2654
 QY 1261 GGTTTACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Db 2653 GGTTTACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2594
 QY 1321 GCTGGCGGAGTGGCTTTTCGCGGACTGTTGGATGTTGCTTCCGCAAGAGATTCGCTCAA 1380
 Db 2593 GCTGGCGGAGTGGCTTTTCGCGGACTGTTGGATGTTGCTTCCGCAAGAGATTCGCTCAA 2534
 QY 1381 GCGAATGGGACACGCAATAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1440
 Db 2533 GCGAATGGGACACGCAATAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 2474
 QY 1441 GAAGAAGACGCTGTACGACA ----- 1460
 Db 2473 GAAGAAGACGCTGTACGACAAGGTGAGTTCTTACTAACAACATCCATCCATGGGATTTGATTT 2414
 QY 1461 -----AGGCTGGCAAGCTGAAGCATCGTACTTAAGCATGATGCCAG 1501
 Db 2413 TAAATACATCGCATTTTTCAGGCTGGCAAGCTGAAGCATCGTACTTAAGCATGATGCCAG 2354

1502 CAGGATACCTGGCGTCCAGCTCAACCTGAAGAGTATGTTATTTTGGCGGCTTCAATGCTG 1561
1503 CAGGATACCTGGCGTCCAGCTCAACCTGAAGAGTATGTTATTTTGGCGGCTTCAATGCTG 2294
1562 TTGATGATGTTGCGTGTCCACTGACGCTGGGTGACCAAGCAATGCTACTCCAGTCCCTCC 1621
2293 TTGATGATGTTGCGTGTCCACTGACGCTGGGTGACCAAGCAATGCTACTCCAGTCCCTCC 2234
1622 ATTGCTTGGCTTTCCCAACAGTCAAGATG----- 1652
2233 ATTGCTTGGCTTTCCCAACAGTCAAGATGCGGTGAAGCACCAGTATCATCTGTTCTT 2174
1653 -----GATCCCGCAACATTTTATAGACGATTTTCAAGAG 1684
2173 TTAAACCTATCTAACCTTTTCTCACATAGTCCCGCAACATTTTATAGACGATTTCAAGAG 2114
1685 GCTTACTACTGCTTTCCGAGAACATCTCCGATGATGCTCGGTATGTTCTTGGTGGAT 1744
2113 GCTTACTACTGCTTTCCGAGAACATCTCCGATGATGCTCGGTATGTTCTTGGTGGAT 2054
1745 TACGATACCATAGTCCGGAATGCAACAGACGAGCTAGTGGGATATAATACGTGG 1804
2053 TACGATACCATAGTCCGGAATGCAACAGACGAGCTAGTGGGATATAATACGTGG 1994
1805 AACATAGTACATAGTCCGGAATGCAACAGACGAGCTAGTGGGATATAATACGTGG 1864
1993 AACATAGTACATAGTCCGGAATGCAACAGACGAGCTAGTGGGATATAATACGTGG 1934
1865 GAAATATGATCTCTTGAAGTACGATGCTTTTGGTATGCTTTTGGGCTGATGCTGGC 1924
1933 GAAATATGATCTCTTGAAGTACGATGCTTTTGGTATGCTTTTGGGCTGATGCTGGC 1874
1925 TATTCGCGATGATATCAACAAAGTCTCTGTTGGATGCTCGAATGCTGAGGAGAGCAT 1984
1873 TATTCGCGATGATATCAACAAAGTCTCTGTTGGATGCTCGAATGCTGAGGAGAGCAT 1814
1985 CCRAAGGACATTAAGGAAGCGATTAATTTACCAACGCGGTGAATTCAGGCTAGATGCC 2044
1813 CCCAAGGACATTAAGGAAGCGATTAATTTACCAACGCGGTGAATTCAGGCTAGATGCC 1754
2045 GAAGTGTCTCGGCGCTCTCACTGCTGTTATGATGATGATGATGATGATGATGATGATG 2104
1753 GAAGTGTCTCGGCGCTCTCACTGCTGTTATGATGATGATGATGATGATGATGATGATG 1694
2105 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2121
1693 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1634
2122 -----CAGAGTCCATCTGATATGATGATGATGATGATGATGATGATGATGATGATG 2171
1633 CTGCTGTTTTCAGAGGTCCTATCTGATATGATGATGATGATGATGATGATGATGATGATG 1574
2172 AGGACTTCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 2231
1573 AGGACTTCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 1514
2232 TCTATAGGTTGAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2291
1513 TCTATAGGTTGAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1454
2292 CGATTCCTCCAGCAACATCTCAATTCGAGAA----- 2321
1453 CGATTCCTCCAGCAACATCTCAATTCGAGAAAGGATGATGATGATGATGATGATGATGATG 1394
2322 -----AGAACTTAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 2357
1393 TTCAATTAACATCTGATCTTTCGAGAACTCTAAGCGCTCGCAAGGCTACATACGAAACC 1334
2358 GACCGGTTGTTTAAAGGAAACCGAACCTTGAATTAACCCAAAAAATAA 2412
1333 GACCGGTTGTTTAAAGGAAACCGAACCTTGAATTAACCCAAAAAATAA 1279

RESULT 5
AC007853/c
LOCUS
DEFINITION
AC007853
AC007853
AC007853.4 GI:13129410
HTG.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 162321)
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,E., Doyle,C., Drenek,D., Farfan,D.,
Fierera,S., Frise,E., Galle,R.F., Gaig,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwan,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nuncio,J.,
Paciel,J., Paragay,V., Park,S., Patel,S., Pfeiffer,S.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 96B-96C
Unpublished
2 (bases 1 to 162321)
Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paciel,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 27, 2001 this sequence version replaced gi:5670481.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgp@fruitfly.berkeley.edu.
Location/Qualifiers
1..162321
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="96B-96C"
/clone="BACR3102 (D766)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"
FEATURES
source
Query Match 88.8%; Score 2147.4; DB 3; Length 162921;
Best Local Similarity 91.5%; Pred No 0;
Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;
QY 1 TCTAAGCGAAGATGCTGCTTCATTCAGATCGGTTAATATTCAGATTCAGTGGCTG 60
|||||

Db 108551 TCTAAGCGAAGATGTGCTGCTCATTTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG 108492
Qy 61 GAATTTGGGACATGAATTCGAGCGCGGAGATGCTGAACAGCAGGTGGCTGGCTACAGCAG 120
Db 108491 GAATTTGGGACATGAATTCGAGCGCGGAGATGCTGAACAGCAGGTGGCTGGCTACAGCAG 108432
Qy 121 CCTAATCACCTTGGCCATCTGCTAATTCGCTGCTGGCTGGCGGATTTTCTCTCGCTCTT 180
Db 108431 CCTAATCACCTTGGCCATCTGCTAATTCGCTGCTGGCTGGCGGATTTTCTCTCGCTCTT 108372
Qy 181 GCGCTTCATCGTTTCGAGTCGATTAATTCATGATGTTTATCGTGGTTCATACCGGCG 240
Db 108371 GCGCTTCATCGTTTCGAGTCGATTAATTCATGATGTTTATCGTGGTTCATACCGGCG 108312
Qy 241 CACCGCTACATGCTGCGAATGTTGGTGAACAATTTCTCAACTGTTTCGAGCGCGC 300
Db 108311 CACCGCTACATGCTGCGAATGTTGGTGAACAATTTCTCAACTGTTTCGAGCGCGC 108252
Qy 301 ATGCTATCCGCTGGCGAGATTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 360
Db 108251 ATGCTATCCGCTGGCGAGATTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 108192
Qy 361 CGCGGGAATCCATTTGGCTGCTGACGCTACTCAACATACCGGTCCCATATTCGTGACATCTG 420
Db 108191 CGCGGGAATCCATTTGGCTGCTGACGCTACTCAACATACCGGTCCCATATTCGTGACATCTG 108132
Qy 421 GGTGTTCTGGCGCGATCTTCTAGTGGCTGACCTCCATCTCCATCTACCTGCTGACCAA 480
Db 108131 GGTGTTCTGGCGCGATCTTCTAGTGGCTGACCTCCATCTCCATCTACCTGCTGACCAA 108072
Qy 481 GGAGCTGTGCTGGCGCGCGCGCTCTTTCGCGCGAGCTTCATCGCATCTGCTGCTGG 540
Db 108071 GGAGCTGTGCTGGCGCGCGCGCTCTTTCGCGCGAGCTTCATCGCATCTGCTGCTGG 108012
Qy 541 CTACATCAGTAGTGGCTGGCTGGATCGTACGATACGAGGCGATTCGATATTCGCT 600
Db 108011 CTACATCAGTAGTGGCTGGATCGTACGATACGAGGCGATTCGATATTCGCT 107952
Qy 601 GCAGTTCACTCTTCTGCTGGCTGGCTCAGTGAAGACTGATCCGCTGTTCTGCTGGC 660
Db 107951 GCAGTTCACTCTTCTGCTGGCTGGCTCAGTGAAGACTGATCCGCTGTTCTGCTGGC 107892
Qy 661 CGAGCGCTTCTCTTCTTCTACATGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 720
Db 107891 CGAGCGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 107832
Qy 721 CAACCTGATACCCCTGACGCTCTTCTGCTGCTCATTATGGCAGGCTACTGCGCGCT 780
Db 107831 CAACCTGATACCCCTGACGCTCTTCTGCTGCTCATTATGGCAGGCTACTGCGCGCT 107772
Qy 781 GCTACACGCTACAGCAGCTTCTACATCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 840
Db 107771 GCTACACGCTACAGCAGCTTCTACATCTGGCTGGCTGGCTGGCTGGCTGGCTGG 107712
Qy 841 CGTGGGATCCAAACGATACGACAGTGAACACATGCTGGCTGGCTGGCTGGCTGGCTGG 900
Db 107711 CGTGGGATCCAAACGATACGACAGTGAACACATGCTGGCTGGCTGGCTGGCTGG 107652
Qy 901 CCTTATGGCTGGCGCTGGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 960
Db 107651 CCTTATGGCTGGCGCTGGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 107592
Qy 961 GAAGCTGTTTCATGCTGGCGGATTTGCTGGCGGCTGGCTGGCTGGCTGGCTGGCTGG 1020
Db 107591 GAAGCTGTTTCATGCTGGCGGATTTGCTGGCGGCTGGCTGGCTGGCTGGCTGGCTGG 107532
Qy 1021 GCTCACCATCTGGCGGCTGGCGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1080
Db 107531 GCTCACCATCTGGCGGCTGGCGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 107472
Qy 1081 TGGCTACGCAAGATCCACATTCCTCATCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1140
Db 107471 TGGCTACGCAAGATCCACATTCCTCATCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 107412

Qy 1141 TTGGTCTCGCTTCTTCTTCTGATCTGCAATCTGCTGGCTGGCTGGCTGGCTGGCTGGCTG 1200
Db 107411 TTGGTCTCGCTTCTTCTTCTTCTGATCTGCAATCTGCTGGCTGGCTGGCTGGCTGGCTG 107352
Qy 1201 GTACTGCTCAAGCAGATCAACGACGAGCGGTTTCTGCTGGCTGGCTGGCTGGCTGGCTG 1260
Db 107351 GTACTGCTCAAGCAGATCAACGACGAGCGGTTTCTGCTGGCTGGCTGGCTGGCTGGCTG 107292
Qy 1261 GGTTTACTTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 1320
Db 107291 GGTTTACTTCTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTG 107232
Qy 1321 GCTGCGGAGTGGCTTTTCTGGGACTGTTGGATGTTGTTCTGCTCAAGAGATTCGCTCAA 1380
Db 107231 GCTGCGGAGTGGCTTTTCTGGGACTGTTGGATGTTGTTCTGCTCAAGAGATTCGCTCAA 107172
Qy 1381 GCGAATGGCAGCAGCATTAAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
Db 107171 GCGAATGGCAGCAGCATTAAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 107112
Qy 1441 GAAGAAGCGCTGTACGACA----- 1460
Db 107111 GAAGAAGCGCTGTACGACAAGGTGAGTTCTTACTAACACACATCCATGGGATTTGATTT 107052
Qy 1461 ----- 1501
Db 107051 TAAATATACATGCAATTCGAGCTGGCAAGCTGAAGCATGTTGTAAGCATGTTAAGCATGTTGCCAG 106992
Qy 1502 CAGGATCTGGCTCAGTCCCACTGAGAGATTTGTTATTTTGGCGCTTCTTAATGCTG 1561
Db 106991 CAGGATCTGGCTCAGTCCCACTGAGAGATTTGTTATTTTGGCGCTTCTTAATGCTG 106932
Qy 1562 TTGATGATCTGGCTCCACTGACAGTGGCTGACAGCAATGCTTACTCCAGTCCCTCC 1621
Db 106931 TTGATGATCTGGCTCCACTGACAGTGGCTGACAGCAATGCTTACTCCAGTCCCTCC 106872
Qy 1622 ATTGCTTGGCTTCCCAACAGTCAAGAT----- 1652
Db 106871 ATTGCTTGGCTTCCCAACAGTCAAGATGGGTGAAGCAAGTAGCATCATGCTTCT 106812
Qy 1653 ----- 1684
Db 106811 TTAACCTATCTAACCTTTTCTCCACTAGATCCCGCAACATTTTGAAGCATTTACAGAG 106752
Qy 1685 GCTTACTAGCTTTCGCAACACATGCTCGATGATGCTCGCTTATGCTTGGTGGGAT 1744
Db 106751 GCTTACTAGCTTTCGCAACACATGCTCGATGATGCTCGCTTATGCTTGGTGGGAT 106692
Qy 1745 TACGATACCGATAGCGGGAATGGCAACAGACAGCAGCTAGTGGATAATAATACGTGG 1804
Db 106691 TACGATACCGATAGCGGGAATGGCAACAGACAGCAGCTAGTGGATAATAATACGTGG 106632
Qy 1805 AACATAGTCAATAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 1864
Db 106631 AACATAGTCAATAGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTAC 106572
Qy 1865 GAAATATGACATCTTTCAGTGGCTACGCTTTTGGTGGATCTTGGCGGTGATCGGC 1924
Db 106571 GAAATATGACATCTTTCAGTGGCTACGCTTTTGGTGGATCTTGGCGGTGATCGGC 106512
Qy 1925 TATTCTGGCGATGATCAACAGATTCCTGTGGATGTTCCGAATGCTGAGGAGAGCAT 1984
Db 106511 TATTCTGGCGATGATCAACAGATTCCTGTGGATGTTCCGAATGCTGAGGAGAGCAT 106452
Qy 1985 CCCAAGGACATTAAGGAAGCGATTTACTTTTACCGACCGGTGAATTCAGGCTAGATGCC 2044
Db 106451 CCCAAGGACATTAAGGAAGCGATTTACTTTTACCGACCGGTGAATTCAGGCTAGATGCC 106392
Qy 2045 GAAGTGTCTCGGCGCTTCTCACTGCTGCTTATGTAATAAATTAAGCTACTACAGATTCGG 2104
Db 106391 GAAGTGTCTCGGCGCTTCTCACTGCTGCTTATGTAATAAATTAAGCTACTACAGATTCGG 106332

```

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.

Location/Qualifiers
1. 181132
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain_1;" cn bw sp"
/db_xref="taxon/7227"
/chromosomes="3R"
/map="96B-96B"
/clone="BACR03115 (D765)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"

FEATURES
source
Query Match 88.8%; Score 2147.4; DB 3; Length 181132;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;
ORIGIN

```

| | | | |
|--------|----|--|--------|
| 171883 | Db | TCTAACGGAAGAAATGGTGGTTCGATTTTCAGATCGGTTATAATTTTCGAGTTTACTGGCTG | 171883 |
| 61 | Qy | GAATTGGGACATGAATTCGGACGCCGAAGATGCTGAACCAAGTGGCTGGCTACAGCAG | 120 |
| 171823 | Db | GAATTGGGACATGAATTCGGACGCCGAAGATGCTGAACCAAGTGGCTGGCTACAGCAG | 171764 |
| 121 | Qy | CTTAATCACCTTCGGCATCCTGCTAATCGCCTGGCTGGCGCGAATTTCCCTCTCGGCTCTT | 180 |
| 171763 | Db | CTTAATCACCTTCGGCATCCTGCTAATCGCCTGGCTGGCGCGAATTTCCCTCTCGGCTCTT | 171704 |
| 181 | Qy | CGCGGTCACTCCGTTTCGAGTCGATTAATCCATGAGTTTGATCCGTGGTTCAACTACCGGCG | 240 |
| 171703 | Db | CGCGGTCACTCCGTTTCGAGTCGATTAATCCATGAGTTTGATCCGTGGTTCAACTACCGGCG | 17164 |
| 241 | Qy | CACCGCTACATGATGTCAGATGTTGTCACAACTTCTCAACTGGTTCGACGAGCGGCG | 300 |
| 171643 | Db | CACCGCTACATGATGTCAGATGTTGTCACAACTTCTCAACTGGTTCGACGAGCGGCG | 17158 |
| 301 | Qy | ATGGTATCCGCTCGCAGGATTTGGCGGGTACCGTCTATCCGCGCTGATGATTACGTC | 360 |
| 171583 | Db | ATGGTATCCGCTCGCAGGATTTGGCGGGTACCGTCTATCCGCGCTGATGATTACGTC | 17152 |
| 361 | Qy | CGCGCGGAATCCAAATGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG | 420 |
| 171523 | Db | CGCGCGGAATCCAAATGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG | 17146 |
| 421 | Qy | CGTGTTCCTGGCGCGATCTTTCAGTGGCCTGACCTCCATCTCCACCTACCTGCTGACCAA | 480 |
| 171463 | Db | CGTGTTCCTGGCGCGATCTTTCAGTGGCCTGACCTCCATCTCCACCTACCTGCTGACCAA | 17140 |
| 481 | Qy | GGAGCTGTGGTTCGGCGGGCGCGGCTCTTCGCGCGCGAGTTCATCGCCATCGGTGCGTGG | 540 |
| 171403 | Db | GGAGCTGTGGTTCGGCGGGCGCGGCTCTTCGCGCGCGAGTTCATCGCCATCGGTGCGTGG | 17134 |
| 541 | Qy | CTACATCAGTAGTTCGGTGGCTGGATCGTACGATAACGAGGGCATTCGCCATATTCGCCCT | 600 |
| 171343 | Db | CTACATCAGTAGTTCGGTGGCTGGATCGTACGATAACGAGGGCATTCGCCATATTCGCCCT | 17128 |
| 601 | Qy | CGAGTTCACTTACTTCTGTGGGTGGCTCAGTGAAGACTGGATCCGCTGTTCTGGTCGGC | 660 |
| 171283 | Db | CGAGTTCACTTACTTCTGTGGGTGGCTCAGTGAAGACTGGATCCGCTGTTCTGGTCGGC | 171224 |
| 661 | Qy | CGGAGCGCGCTTGTCTCACTTCTACATGGTGTCCGCTGGGGTGGCTACGTTGTTCATCAT | 720 |

| | | | |
|----|--------|---|--------|
| Db | 171223 | CGAGCGCGCTTTGTCCTACTTCTACATGGTGTCCGCTCGGGTGGCTACGTGTTCAATCAT | 171166 |
| QY | 721 | CAACCTGATACCCCTGCACTGCTTTCTGACTGCTCATTTATGGGAGAGTACTCGCCGCGCT | 780 |
| Db | 171163 | CAACCTGATACCCCTGCACTGCTTTCTGACTGCTCATTTATGGGAGAGTACTCGCCGCGCT | 171104 |
| QY | 781 | GCTGACCACTACAGACCTTCTACATCCCTGGGAGTCTGTTCTCCATCGAGTCCCTT | 840 |
| Db | 171103 | GCTGACCACTACAGACCTTCTACATCCCTGGGAGTCTGTTCTCCATCGAGTCCCTT | 171044 |
| QY | 841 | CGTGGGATTCGAACCGATACGACACAGTGAAACATATGGCTGCGCTCGGAGTGTGTGTCT | 900 |
| Db | 171043 | CGTGGGATTCGAACCGATACGACACAGTGAAACATATGGCTGCGCTCGGAGTGTGTGTCT | 170984 |
| QY | 901 | CCTTATGGCGTGGCCACTTGTGGCCATTTTGCACTGCGTCTGTCGCGCAACGAGTTCGG | 960 |
| Db | 170983 | CCTTATGGCGTGGCCACTTGTGGCCATTTTGCACTGCGTCTGTCGCGCAACGAGTTCGG | 170924 |
| QY | 961 | GAAGCTGTTATGCTGCGCGGATTCGTTGGGGTTGGCGTCTTTGTGGCCGTGTGTGT | 1020 |
| Db | 170923 | GAAGCTGTTATGCTGCGCGGATTCGTTGGGGTTGGCGTCTTTGTGGCCGTGTGTGT | 170864 |
| QY | 1021 | GCTCACCAATGCTGGCGGTTTGGCCCGTGGAGTGACGCTTCTACTCGTGTGGAGTAC | 1080 |
| Db | 170863 | GCTCACCAATGCTGGCGGTTTGGCCCGTGGAGTGACGCTTCTACTCGTGTGGAGTAC | 170804 |
| QY | 1081 | TGGCTACGCCAAGATPCCATATCCCATCATTTGCATCCGTCTCGAGCATACGCCACCAAC | 1140 |
| Db | 170803 | TGGCTACGCCAAGATPCCATATCCCATCATTTGCATCCGTCTCGAGCATACGCCACCAAC | 170744 |
| QY | 1141 | TTGGTTCTCGTCTCTTTTGATCTGCAATCCTGTGTGCGCTTCCCACTGGGAGTGTG | 1200 |
| Db | 170743 | TTGGTTCTCGTCTCTTTTGATCTGCAATCCTGTGTGCGCTTCCCACTGGGAGTGTG | 170684 |
| QY | 1201 | GTACTGCATTAAGCAGATCAACGACGAGCGGTTTTCTGGTGTCTGTACGCCATCAGTGC | 1260 |
| Db | 170683 | GTACTGCATTAAGCAGATCAACGACGAGCGGTTTTCTGGTGTCTGTACGCCATCAGTGC | 170624 |
| QY | 1261 | GGTTTACTTTCGTGTGTGATGTGGTGGCTTTGATGTTGACCTCACGCCGTTGGTGTGCAT | 1320 |
| Db | 170623 | GGTTTACTTTCGTGTGTGATGTGGTGGCTTTGATGTTGACCTCACGCCGTTGGTGTGCAT | 170564 |
| QY | 1321 | GCTGGCCGAGTGGCTTTTTCGGAGCTGTGGATGTGTTCTCTCCAGAGAGATTGCTCTAA | 1380 |
| Db | 170563 | GCTGGCCGAGTGGCTTTTTCGGAGCTGTGGATGTGTTCTCTCCAGAGAGATTGCTCTAA | 170504 |
| QY | 1381 | GGGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTTCATTGA | 1440 |
| Db | 170503 | GGGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTTCATTGA | 170444 |
| QY | 1441 | GAAGAGACGCTGTAGACA----- | 1460 |
| Db | 170443 | GAAGAGACGCTGTAGACAAGGTGATCTTACTAACACATCCATGGGATTTGATTT | 170384 |
| QY | 1461 | -----ASGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG | 1501 |
| Db | 170383 | TAAATATACATGCAATTCGAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG | 170324 |
| QY | 1502 | CAGGATCTGGGCTCAGCTCCAACTGAAGATATTGTTATTTTGGCCGTTCTAATGCTG | 1561 |
| Db | 170323 | CAGGATCTGGGCTCAGCTCCAACTGAAGATATTGTTATTTTGGCCGTTCTAATGCTG | 170264 |
| QY | 1562 | TTGATGATTTGGCTGTCCATCGACGTGGGTGACCAAGCAATGCCTACTCCAGTCCCTCC | 1621 |
| Db | 170263 | TTGATGATTTGGCTGTCCATCGACGTGGGTGACCAAGCAATGCCTACTCCAGTCCCTCC | 170204 |
| QY | 1622 | ATTGCTTGGCTTTCCCAACAGTCAAGTG----- | 1652 |
| Db | 170203 | ATTGCTTGGCTTTCCCAACAGTCAAGTGGGTAGACCAAGTAGCATATGCTTCT | 170144 |
| QY | 1653 | -----GATCCCGCAACATTTTAGACATTTCCAGAG | 1684 |
| Db | 170143 | TTAAACCTATTAACCTTTTCTCCACTAGATCCCGCAACATTTAGACATTTCCAGAG | 170084 |

| | | | |
|----|--------|---|--------|
| Qy | 1685 | GCTTACTACTGGCTTTTCGAGAACTCCGATGATGCTGCGTTATGTCTTTGGTGGAT | 1744 |
| Db | 170083 | GCTTACTACTGGCTTTTCGAGAACTCCGATGATGCTGCGTTATGTCTTTGGTGGAT | 170024 |
| Qy | 1745 | TACGGATACCAAGATAGCGGNAATGGCAAAACAGAACGACCTAGTGGATAATAATACGTGG | 1804 |
| Db | 170023 | TACGGATACCAAGATAGCGGNAATGGCAAAACAGAACGACCTAGTGGATAATAATACGTGG | 169964 |
| Qy | 1805 | AACAAATAGTCACATAGCGCTGGTTTGGCAAGGCAATGCTTCAACGAGGAGAAGTCCTAC | 1864 |
| Db | 169963 | AACAAATAGTCACATAGCGCTGGTTTGGCAAGGCAATGCTTCAACGAGGAGAAGTCCTAC | 169904 |
| Qy | 1865 | GAATAATATGACATCTCTTTCGATCGGACTAGCTTTTGGTGATCTTTGGCGGCTGATCGGC | 1924 |
| Db | 169903 | GAATAATATGACATCTCTTTCGATCGGACTAGCTTTTGGTGATCTTTGGCGGCTGATCGGC | 169844 |
| Qy | 1925 | TATTCCTGGCGATGATATCAACAAGTTCCTGTGGATGGTCGAATTCCTCAGGAGAGAGAT | 1984 |
| Db | 169843 | TATTCCTGGCGATGATATCAACAAGTTCCTGTGGATGGTCGAATTCCTCAGGAGAGAGAT | 169784 |
| Qy | 1985 | CCCAAGSACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGTTAGATGCC | 2044 |
| Db | 169783 | CCCAAGSACATTAAGGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGTTAGATGCC | 169724 |
| Qy | 2045 | GAAGTGTCTCGGCGCTTCCTCACTGCCTTATGTACAAATTAAGCTACTACAGATTCGG | 2104 |
| Db | 169723 | GAAGTGTCTCGGCGCTTCCTCACTGCCTTATGTACAAATTAAGCTACTACAGATTCGG | 169664 |
| Qy | 2105 | GAATTGAAGTTGGACTA----- | 2121 |
| Db | 169663 | GAATTGAAGTTGGACTACAGSTAAGCGGAACAATTTCTCAGGTAGCGATCCACACTAA | 169604 |
| Qy | 2122 | -----CAGAGGTCCATCTGGATATGATCGCACAGTAACGCGCTCATTTGGGAATA | 2171 |
| Db | 169603 | CTGCGCTGTTTCAGAGGTCCATCTGGATATGATCGCACAGTAACGCGCTCATTTGGGAATA | 169544 |
| Qy | 2172 | AGGACTTCGATCTGACCTACTCTGAGGAGGSCCTACACCAGAACACTGGCTTGTTCGCA | 2231 |
| Db | 169543 | AGGACTTCGATCTGACCTACTCTGAGGAGGSCCTACACCAGAACACTGGCTTGTTCGCA | 169484 |
| Qy | 2232 | TCTATAGGGTGAAGAAGCCGCGATGATTCAATAGACCATCTACGTAGAGACCAAGGAGAA | 2291 |
| Db | 169483 | TCTATAGGGTGAAGAAGCCGCGATGATTCAATAGACCATCTACGTAGAGACCAAGGAGAA | 169424 |
| Qy | 2292 | CGATTCCTCCAGCAAACTTCATTTCCGAA----- | 2321 |
| Db | 169423 | CGATTCCTCCAGCAAACTTCATTTCCGAAAGAGGTAGTGATCCGTCAGCTCCCAATGAAC | 169364 |
| Qy | 2322 | -----AGAACTCTTAAGCGTCGCAAGGCTACATCCGAAACC | 2357 |
| Db | 169363 | TTTCATTAACATCTGATTCITTCGAGAACTCTTAAGCGTCGCAAGGCTACATCCGAAACC | 169304 |
| Qy | 2358 | GAACGGTGTGTTAAGGGAAGAACGAACCTTGAATAAACCMAAAAAAAAAA | 2412 |
| Db | 169303 | GAACGGTGTGTTAAGGGAAGAACGAACCTTGAATAAACCMAAAAAAAAAA | 169249 |

| | |
|------------|-----------|
| RESULT 7 | REFERENCE |
| AE003750/C | AUTHORS |
| LOCUS | |
| DEFINITION | |
| ACCESSION | |
| VERSION | |
| KEYWORDS | |
| SOURCE | |
| ORGANISM | |

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Workman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abil, J.F., Agbayan, A., An, H.J., Andrews-Pfannkuch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, B.M., Beeson, K.Y., Bence, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Bottler, P., Burris, K.C., Busam, J.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, R., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Diez, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S., Fleischmann, W., Fosler, C., Gabriellian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.I., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibgweg, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, B., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 10731132

2 (bases 1 to 227219)
 Celnikier, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., Hodgson, A., George, R.A., Hoskins, R.A., Lavery, T., Muzny, D.M., Nelson, C.R., Pacle, J.M., Park, S., Pfeiffer, B.D., Richards, S., Sodergren, E.J., Svirskas, R., Tabor, P.E., Wan, K., Stapleton, M., Sutton, G.G., Venter, C., Weinstein, G., Scherer, S.E., Myers, E.W., Gibbs, R.A., and Rubin, G.M.

Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence
 Genome Biol. 3 (12), RESEARCH0079 (2002)
 2426065
 12537568

3 (bases 1 to 227219)
 Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradecky, P., Huang, Y., Kaminker, J.S., Milburn, G.H., Prochuk, S.E., Smith, C.D., Tupy, J.L., Whitfield, E.J., Bayraktaroglu, L., Berman, B.P., Bettencourt, B.R., Celnikier, S.E., de Grey, A.D., Drysdale, R.A., Harris, N.I., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M., and Lewis, S.E.

Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review
 Genome Biol. 3 (12), RESEARCH0083 (2002)
 2242609
 12537572

4 (bases 1 to 227219)
 Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M., and Celnikier, S.E.

The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective
 Genome Biol. 3 (12), RESEARCH0084 (2002)

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

22426070
 12537573
 5 (bases 1 to 227219)
 Adams, M.D., Celnikier, S.E., Gibbs, R.A., Rubin, G.M., and Venter, C.J.

Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 6 (bases 1 to 227219)

FlyBase
 Direct Submission
 Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA
 7 (bases 1 to 227219)

FlyBase
 Direct Submission
 Submitted (10-MAR-2004) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, USA
 On or before Sep 18, 2002 this sequence version replaced
 gi:10729289, gi:7301241.

Location/Qualifiers
 1. .227219
 /organism="Drosophila melanogaster"
 /mol_type="Genomic DNA"
 /db_xref="taxon:7227"
 /chromosome="3R"
 complement(149..6378)
 /gene="ssh"
 /locus_tag="CG6238"
 /map="56810-96811"
 /db_xref="FLYBASE:FBgn0029157"
 complement(join(149..3227,3313..3582,3649..3843,3907..3972,4265..4656,4722..4967,5864..6378))
 /gene="ssh"
 /locus_tag="CG6238"
 /product="CG6238-RA"
 /db_xref="FLYBASE:FBgn0029157"
 complement(join(149..3227,3313..3582,3649..3843,3907..3972,4265..4656,4722..4967,5864..6378))
 /gene="ssh"
 /locus_tag="CG6238"
 /product="CG6238-RB"
 /db_xref="FLYBASE:FBgn0029157"
 complement(join(875..3227,3313..3582,3649..3843,3907..3972,4265..4656,4722..4967,5864..6378))
 /codon_start=1
 /product="CG6238-PA"
 /protein_id="AAF56372.3"
 /db_xref="GI:23172230"
 /translation="MALVTQSPSPVAGSCNSDGESEDDGNSKGNDRSECFPAAGK
 VALVALKDPIPTQSVRRSLTSTSRNSSTQSNNSDIQLHLSQMFYLLQREDTLKMA
 TKLQSRQNRTRIVIAASRCCSGTSDRRHRIMRHHSVKVGSAGCTKSTSPAVT
 QROSVQQTAEASCKDADKENATAAGNKTSCWEECLGIDCNRTHGLVY
 PILADTTIHLDDGSPSVKYKTHIFKPVSVQAMGALQTLHKVSKARENPFYASG
 PSHDWLSYRERIESDQSCLENNAMDALSRRPPSDAIRKPKKEETESVIMKML
 KAINMSVLDLETSKYIRGLREILDMDLGEYKFFDAEMLVILQMDATPKIPEHY
 LGSWNASNLEEIQNGVRHILNVTREIDNFFEFYFNVRVYDEDTLLKYDDT
 FRYTRAKAEGSKVLVHCNKGVSRSASVIAYAMKAYQWFEQALHKKRRSCIKEN
 CNFLNQLTSGMLDMANKKELQKRSKSTNLKSTADARLLPGSEPTLIQALNOAKS
 KSTGEAGVTPGDEEDGSRHRSIAQKSORRMVRSSTSPKQTAVVTQKQOSME
 NLTPERSVAREPKNRPFGNGSNYSVTQVLIQKHTPLSVTRTVDHLEARADQL
 POQPVNTSLTKLTQTSHLGKSVSGSSGNDISRRSSCSDFSSQVDSVFAKDEGK
 RQRKTHSWTESLPGSGVITDPTPOQKQSNAILRPGRTQREDFSRHSGSGEDV
 RCLQPTSSGSGYSDNRNTTAFEGVIOQLKSSNVCNIEGVAVAPVIGVGEQTV
 KRTKQLEESTSLKRCQESQELLLEAVDAGQRCPSLYRSASHSRQRQPLRCN
 SEELMHTSDIENKNTTPGHEATVLRVQCGTQTLADQRCNQLKONFEAKGVVG
 TGGCGGGTAAAGSGSSTIATSVAVAGCANGPKKTTSHQSLPSPVAQHNVSA
 ASNSNSASNSSDSVSENCNRVKGIVHKYQTTTSCSSQAAATPVACALPPAGTA
 SNLSHSCSTTTTTSFKQRPVLSYDRSLNSRSHDYSRPPFPAPVSRVSHSNHQOQ

| | | | | | |
|-----|---|---|--------|---|-----|
| CDS | OOQOLQPEHQOQVOIRLAHQKTHPLSELSPAKTGFNTAAAYNTM" complement(join(875..3227,3313)..3582,3649..3843, 3907..3972,4265..4656,4722..4967,5036..5095)) /gene="ssh" /locus_tag="CG6238" /codon_start=1 /product="CG6238-PB" /protein_id="AANI4027.1" /db_xref="GI:2317231" /db_xref="FLYBASE:FBgn0029157" /translation="MSKQVCSLEDDLLIFSLTPRDSSEDDNSKGNDRSCFPAGK GTALVALKDIPPTQSERLSTDSSTOSNSDIQLHLQSMFYLLQREDTLKM AVKLESNRNRYLYIASRSCRSCTSDRRHRIRNRHSVYKVSAGCTSTSPAVP TORQLSEQATASASCKDRTAKENATAAGDNKNTSGMEESCLIGDCNERTITGLV VPIADTTHLDGPGFSVKVYEKTHFKPVSQVAMWSALQTLHKVSKKARNFYAS GFSHOWLSSEYRRTESDQSCNLEWAMDALESRPPSPDAIRNKPKEETESVLMK LKAIMSVLDIDVENTSKYIRGLEILLDMOLGEYKSIDAEMLVILQMDAPKPIEHV YLGEWNASLLEHCKMGVRIHLNVTREIDNFFGTFEIRNVRVIDEKINLLKWD TFRIIRAKAGSKVLVHCKMGVRSASVVIAYAMKAYQWFEQALDEHKVRRSCIKP NNFLNQLETYSGLDMKNKELQRSKSETNLKSTKDARLLPSEFPPLQALNQAK SKSTGEAGTVPDGEEDGSRHRSJAQSRMRRSSSTSPKQTAVTVKQOSQM ENLTPERSVAEPPNMFPGSNGENTSVQNLHIQKHTPLSVTRHDLAHRADQ LPQPVWTSJTKLITQTHLKGVSQSSGNIDSRDSSCDVSFSDVDFAKDEGE KROKRKTHSTESLPGSGGIVLDPQKQKQSNALRLPRGRORLEPSRSHASWGSJ NRCLPORTSGSYDSNRNNTALPEGVLODKRBSNQNVIQVAVPEPIVGEGT VKRTQKLESTSLKRCQESQELLLEAVDAQRCRPSLYRSASASAPROQPLRC NSBELMHSIDENKNTTPGDHEATVILRVGGTLDQRIQSGNVQILKONFEAKGVV GTGGGGGGTASGSSSTIATSVAGACANPKGKTTSHQSLPSPVAQHANVS AASNSASNSDSSSVENCDNRNKLGHVKTQTHSSQAAATPTVACAPLPAGT ASNTSHSCSTSTTSFKRPLSYDRSLNRSHTDYSRPPPPAPVRYVSHSNVQHQ QOOOOLQPEHQOQVOIRLAHQKTHPLSELSPAKTGFNTAAAYNTM" 6683..8846 /locus_tag="CG13645" /map="96B11-96B11" /db_xref="FLYBASE:FBgn0039254" join(6683..6899,6972..7378,7430..7531,7590..7741, 8330..8600,8709..8846) /locus_tag="CG13645" /product="CG13645-RA" /db_xref="FLYBASE:FBgn0039254" join(6683..6899,6972..7378,7430..7531,7590..7741, 8330..8600,8709..8846) | QY | 421 | CGTGTTCCTGGCGCCGATCTTCAGTGGCCCTGACCTCCATCTCCACCTACCTGCTGACCAA | 480 |
| Ds | 130650 | CGTGTTCCTGGCGCCGATCTTCAGTGGCCCTGACCTCCATCTCCACCTACCTGCTGACCAA | 130591 | | |
| QY | 481 | GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGCCTGG | 540 | | |
| Ds | 130590 | GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGCCTGG | 130531 | | |
| QY | 541 | CTACATCAGTAGTCTGGTGGCTGGATCGATCAATCAAGAGGCAATGCAATTCGCTTC | 600 | | |
| Ds | 130530 | CTACATCAGTAGTCTGGTGGCTGGATCGATCAATCAAGAGGCAATGCAATTCGCTTC | 130471 | | |
| QY | 601 | GCAGTTCACCTACTCTCTGGTGGCTGGATCGATCAATCAAGAGGCAATGCAATTCGCTTC | 660 | | |
| Ds | 130470 | GCAGTTCACCTACTCTCTGGTGGCTGGATCGATCAATCAAGAGGCAATGCAATTCGCTTC | 130411 | | |
| QY | 661 | CGAGCGCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 720 | | |
| Ds | 130410 | CGAGCGCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 130351 | | |
| QY | 721 | CAACCTGATACCCCTGCAGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 780 | | |
| Ds | 130350 | CAACCTGATACCCCTGCAGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 130291 | | |
| QY | 781 | GCTGACAGCTACAGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 840 | | |
| Ds | 130290 | GCTGACAGCTACAGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 130231 | | |
| QY | 841 | CGTGGGATTCCTCAACCGATACGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 900 | | |
| Ds | 130230 | CGTGGGATTCCTCAACCGATACGACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 130171 | | |
| QY | 901 | CCTTATGGCGCTGGCGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 960 | | |
| Ds | 130170 | CCTTATGGCGCTGGCGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 130111 | | |
| QY | 961 | GAAGCTGTTCATCGTGGCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1020 | | |
| Ds | 130110 | GAAGCTGTTCATCGTGGCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 130051 | | |
| QY | 1021 | GCTCAGATCTGGGGGCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1080 | | |
| Ds | 130050 | GCTCAGATCTGGGGGCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 129991 | | |
| QY | 1081 | TGCTACGCAAGATCCACATTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 1140 | | |
| Ds | 129990 | TGCTACGCAAGATCCACATTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 129931 | | |
| QY | 1141 | TTGGTCTCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 1200 | | |
| Ds | 129930 | TTGGTCTCTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 129871 | | |
| QY | 1201 | GTACTCCATCAAGAGATCAACGAGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 1260 | | |
| Ds | 129870 | GTACTCCATCAAGAGATCAACGAGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 129811 | | |
| QY | 1261 | GGTTTACTTCTGCTGGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1320 | | |
| Ds | 129810 | GGTTTACTTCTGCTGGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 129751 | | |
| QY | 1321 | GCTGGCGGAGTGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 1380 | | |
| Ds | 129750 | GCTGGCGGAGTGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT | 129691 | | |
| QY | 1381 | GCGAATGGGCAAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG | 1440 | | |
| Ds | 129690 | GCGAATGGGCAAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG | 129631 | | |
| QY | 1441 | GAAGAGAGCGCTGTACGACA----- | 1460 | | |
| Ds | 129630 | GAAGAGAGCGCTGTACGACA----- | 129571 | | |

| | | | |
|----|--------|--|--------|
| Qy | 1461 | -----AGSGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG | 1501 |
| Db | 129570 | TAAATACATACGCAATTGTCAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG | 129511 |
| Qy | 1502 | CAGGATACCTGGCGTCAGCTCCAACTGAAGAGTATTGTTATTTGGCGCTTCCTAATGCTG | 1561 |
| Db | 129510 | CAGGATACCTGGCGTCAGCTCCAACTGAAGAGTATTGTTATTTGGCGCTTCCTAATGCTG | 129451 |
| Qy | 1562 | TTGATGATGTCGCTGTCACATGCACTGGGCTGACCAAGCAATGCTCTACCTCCAGTCCCTCC | 1621 |
| Db | 129450 | TTGATGATGTCGCTGTCACATGCACTGGGCTGACCAAGCAATGCTCTACCTCCAGTCCCTCC | 129391 |
| Qy | 1622 | ATTGCTCTGGCTTTCCACAACAGTCAAGATG----- | 1652 |
| Db | 129390 | ATTGCTCTGGCTTTCCACAACAGTCAAGATGCGGTAAGCAACCAAGTAGCATCATGCTTCT | 129331 |
| Qy | 1653 | -----GATCCCGCAACATTTTAGACGATTTACAGAG | 1684 |
| Db | 129330 | TTAAACCTATCTAACTTTTCTCCTAGATCCCGCAACATTTTAGACGATTTACAGAG | 129271 |
| Qy | 1685 | GCTTACTACTGCTTTCCAGAACACTCCCATGATGCTCGCGTTATGCTTGGTGGGAT | 1744 |
| Db | 129270 | GCTTACTACTGCTTTCCAGAACACTCCCATGATGCTCGCGTTATGCTTGGTGGGAT | 129211 |
| Qy | 1745 | TACGGATACAGATAGCGGAATGCAACAGAACAGCGCTAGTGGATATATATAGCTGG | 1804 |
| Db | 129210 | TACGGATACAGATAGCGGAATGCAACAGAACAGCGCTAGTGGATATATATAGCTGG | 129151 |
| Qy | 1805 | AACAATAGTCATACGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAACTCTATC | 1864 |
| Db | 129150 | AACAATAGTCATACGCTGGTTGGCAAGCAATGCTTCAACCGAGGAGAACTCTCTAC | 129091 |
| Qy | 1865 | GAATATATCACAATCTCTTGACCTGACATACGTTTGTGTGATCTTTGGCGGTGTATCGGC | 1924 |
| Db | 129090 | GAATATATCACAATCTCTTGACCTGACATACGTTTGTGTGATCTTTGGCGGTGTATCGGC | 129031 |
| Qy | 1925 | TATCTGGCGATGATATCAACAAGTTCCTGTGGATGGTCCGAATGCTGAGGAGAGAGAT | 1984 |
| Db | 129030 | TATCTGGCGATGATATCAACAAGTTCCTGTGGATGGTCCGAATGCTGAGGAGAGAGAT | 128971 |
| Qy | 1985 | CCCAAGCAATTAAGAAAGCGATTACTTTACCACCGCGGTGAATTCAGGGTAGATGCC | 2044 |
| Db | 128970 | CCCAAGCAATTAAGAAAGCGATTACTTTACCACCGCGGTGAATTCAGGGTAGATGCC | 128911 |
| Qy | 2045 | GAAGTGTCTCGGCGCTGCTCAACTGCTCTTATGTACAAATTAAGCTACTACAGATTCCGG | 2104 |
| Db | 128910 | GAAGTGTCTCGGCGCTGCTCAACTGCTCTTATGTACAAATTAAGCTACTACAGATTCCGG | 128851 |
| Qy | 2105 | GAATTCAGGTTCGACTA----- | 2121 |
| Db | 128850 | GAATTCAGGTTCGACTACAGGTAAGCGGAACAATTTCTCTCAGGTAGCATGCACACTAA | 128791 |
| Qy | 2122 | -----CAGAGGTCACTCTGGATATGATCCGACAGTAAAGCGCTCATTTGGGAATA | 2171 |
| Db | 128790 | CTGCCCTGTTTCAGAGGTCCATCTGGATATGATCCGACAGTAAAGCGCTCATTTGGGAATA | 128731 |
| Qy | 2172 | AGGACTTCGATCTGACCTACCTGGAGGAGGCGCTACACAGAACTGGCTTTGTTCCGA | 2231 |
| Db | 128730 | AGGACTTCGATCTGACCTACCTGGAGGAGGCGCTACACAGAACTGGCTTTGTTCCGA | 128671 |
| Qy | 2232 | TCTATAGGTGAAGAAGCGGATGATTTCAATATAGCAATCATCTGAAGACCAAGGAGAGAA | 2291 |
| Db | 128670 | TCTATAGGTGAAGAAGCGGATGATTTCAATATAGCAATCATCTGAAGACCAAGGAGAGAA | 128611 |
| Qy | 2292 | CGATTCTCTCAGCAAACTTCATTTCCGAAA----- | 2321 |
| Db | 128610 | CGATTCTCTCAGCAAACTTCATTTCCGAAAAGGTAGTGTACCTGCACTCCCAATGAAC | 128551 |
| Qy | 2322 | -----NCACTCTAAGCTCCGAGGCTACATACGAAAC | 2357 |
| Db | 128550 | TTCAATTAACAATCTGATCTTTTCAGAACTCTTAAGCGTCCGACAGGGCTCATACGAAAC | 128491 |
| Qy | 2358 | GACCGGTGCTGTGTTAAGGAAACCAACCTTGAATATAACCCAAAAAATAAAAAA | 2412 |

| Db | 128490 | GACCGTGTGTGTTAAGGAAACGAACCTTGAAATAAACCCAAAAAATAA | 128436 |
|-----------------------|--|---|------------------------|
| RESULT 8 | | | |
| AC018145/c | | | |
| LOCUS | AC018145 | 75650 bp DNA linear | HTG 09-DEC-1999 |
| DEFINITION | Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** | | |
| ACCESSION | AC018145 | | |
| VERSION | AC018145.1 | GI:6553046 | |
| KEYWORDS | HTG; HTGS_PHASE2. | | |
| SOURCE | Drosophila melanogaster (fruit fly) | | |
| ORGANISM | Drosophila melanogaster | | |
| | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | |
| | Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; | | |
| | Ephydroidea; Drosophilidae; Drosophila. | | |
| REFERENCE | 1. (bases 1 to 75650) | | |
| AUTHORS | Adams,M. and Venter,J.C. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA | | |
| COMMENT | <p>This sequence was identified as CDM:10213542 by the submitter. For more information on this record e-mail to fly@celera.com.</p> <p>* NOTE: This is a 'working draft' sequence.</p> <p>* This sequence will be replaced</p> <p>* by the finished sequence as soon as it is available and</p> <p>* the accession number will be preserved.</p> | | |
| FEATURES | <p>Location/Qualifiers</p> <p>1..75650</p> <p>/organism="Drosophila melanogaster"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:7227"</p> | | |
| ORIGIN | | | |
| Query Match | 88.2%; | Score 2131.6; | DB 2; Length 75650; |
| Best Local Similarity | 91.4%; | Pred. No. 0; | |
| Matches 2407; | Conservative | 0; Mismatches | 4; Indels 223; Gaps 5; |
| QY | 1 | TCTAACGGAAGATGTCGTTGCATTTTCAGATCGGTTATAATTTTCGAGTTACTGGCTG | 60 |
| DB | 21557 | TCTAACGGAAGATGTCGTTGCATTTTCAGATCGGTTATAATTTTCAGTTACTGGCTG | 21498 |
| QY | 61 | GAATTCGGACATGATCGAGCCGGAAGATGCTGAACGACGAGTGCTGCTACAGCAG | 120 |
| DB | 21497 | GAATTCGGACATGATTCGGACGCCGAAGATGCTGAACGACGAGTGCTGCTACAGCAG | 21438 |
| QY | 121 | CCTAATCAGCTTGCCATCCTCTAATCGCTTGCTGCGCGGATTTTCTCTCGCTCTT | 180 |
| DB | 21437 | CCTAATCAGCTTGCCATCCTCTAATCGCTTGCTGCGCGGATTTTCTCTCGCTCTT | 21378 |
| QY | 181 | CGCGGTATCCGTTTCGAGTCGATTCATCAGTTTGATCGTGCGTTCACCTACCGGC | 240 |
| DB | 21377 | CGCGGTATCCGTTTCGAGTCGATTCATCAGTTTGATCGTGCGTTCACCTACCGGC | 21318 |
| QY | 241 | CACCGCTACATGGTGAGATGGTTGGTACAACTTCCTCAACTCGGTTTCGACGCGCG | 300 |
| DB | 21317 | CACCGCTACATGGTGAGATGGTTGGTACAACTTCCTCAACTCGGTTTCGACGCGCG | 21258 |
| QY | 301 | ATGGTATCCGCTCGGACGATTTGTGGCGGTACCGTCTATCCGSCCTGATGATTACGTC | 360 |
| DB | 21257 | ATGGTATCCGCTCGGACGATTTGTGGCGGTACCGTCTATCCGSCCTGATGATTACGTC | 21198 |
| QY | 361 | CGCGGAATCCATTCGCTGCTGCAGCTACTCAACATACCGGTCCATATTCGTGACATCTG | 420 |
| DB | 21197 | CGCGGAATCCATTCGCTGCTGCAGCTACTCAACATACCGGTCCATATTCGTGACATCTG | 21138 |
| QY | 421 | CGTGTTCCTGGCGCCGATTTTCAGTGGCTGACCTCTCCACTACCTGCTGACCAA | 480 |
| DB | 21137 | CGTGTTCCTGGCGCCGATTTTCAGTGGCTGACCTCTCCACTACCTGCTGACCAA | 21078 |
| QY | 481 | GGAGCTGTGTGTTCGCGGGCGCGGCTCTTTCGCGCGCAGCTTTCATCGCATCTGCTG | 540 |
| DB | 21077 | GGAGCTGTGTGTTCGCGGGCGCGGCTCTTTCGCGCGCAGCTTTCATCGCATCTGCTG | 21018 |

| ORIGIN | | /mol_type="mRNA" /db_xref="taxon:10090" | | 39.7%; Score 960.6; DB 6; Length 2710; Best Local Similarity 64.8%; Pred. No. 1.5e-206; Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4; | |
|--------|------|---|------|--|--|
| QY | 102 | AGTGGCTGGCTACAGCAGCCTAATACACCTTGGCCATCTCTGCTAATCGCTGCTGCGCCG | 161 | | |
| Db | 258 | AGCCGCGCGGGTGGCAGTGGTGTCTCTCTTTCACCATCTCTTCTGCGCTGCGCCG | 317 | | |
| QY | 162 | GAATTTCTCTCTGCTCTTGGCGGTCATCGTTTTCGAGTGGATTATCCATGAGTTTGATC | 221 | | |
| Db | 318 | GCTTCAGCTCGGCTCTTTCGCGCTCATCCGCTTCGAGAGCATCATCCAGAGTTTCAGCC | 377 | | |
| QY | 222 | CGTGGTTCAATACCGGCGCCACCGCTACATGGTGCAGAAATGGTGTGTAACAATCTCTCA | 281 | | |
| Db | 378 | CGTGGTTTAACTATAGATCAACATCATCTCTGCTCATGGATTCTATGAGTTTCTAA | 437 | | |
| QY | 282 | ACTGGTTTCAGAGCGCGCATGGTATCCGCTCGCAGGATTTGTGGCGGTACCGTCTATC | 341 | | |
| Db | 438 | ATTTGGTTTGAAGAGAGATGGTACCCACTTGGGAGAAATAGTGGTGGCACCGTTTACC | 497 | | |
| QY | 342 | CGGCGCTGATGATACGTTCGGCGGAAATCCATTTGGCTGCTGCAGATCTCAACATACCGG | 401 | | |
| Db | 498 | CAGGGTTGATGATAACAGCTGGCGCTTATTTCAATTGGATTAAATACATTTGAACATACAG | 557 | | |
| QY | 402 | TCCATATTCTGACATCTGGGTGTTCTTGGCGCGCATCTTCAGTGGGCTCACCTCCATCT | 461 | | |
| Db | 558 | TTACATAAGAGATGTGTGTATCTTTCCTTGACCAACTTTTAGCGGCTTACATCCATAT | 617 | | |
| QY | 462 | CCACCTACTCTGTGACCAAGAGAGTGGTGGCGGGCGCGGCTCTTCGCGCGCAGCT | 521 | | |
| Db | 618 | CTACCTTCTGCTAACTAGAGAACTGTGGAACTGTGGAACTTAAATACATTTGAACATACAG | 577 | | |
| QY | 522 | TCATCGCATCTGCTGCTGCTACATCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT | 581 | | |
| Db | 678 | TCATTTGCTATCGTACAGAGTACATATCTCGGTGAGTGGCGGATCTTTGATATGAAG | 737 | | |
| QY | 582 | GCAATGGCAATATCGCTCGAGTTTACCTTCTCTGTTGGTGGTGGTGGTGGTGGTGGTGGT | 641 | | |
| Db | 738 | GCAATGGCAATTTTGGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA | 797 | | |
| QY | 642 | GATCGGTGTTCTGGTGGCGCGCGCTTGTGCTCTTCTACATGCTGCTGCTGCTGCTGCTG | 701 | | |
| Db | 798 | GGTCTGTGTTCTGGACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 857 | | |
| QY | 702 | GTGGCTAGCTGTTCATCACTCACTGATACCCCTGACGCTTCTGCTACTGCTCATATGG | 761 | | |
| Db | 858 | GAGGTATGTGTTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT | 917 | | |
| QY | 762 | GCAGTACTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 821 | | |
| Db | 918 | AGAGGTACAGCAAGAGAGTCTACATAGCATATAGCATCTTGTATATTTGGGTTTAAATAT | 977 | | |
| QY | 822 | TCTCATGAGATATCCCTTGTGGATTCGAACCTGATACGACACAGTGAACATAGGCTG | 881 | | |
| Db | 978 | TATCCATGAGATATCTTTTGGGATTTTCAGCAATTCAGCAATTCAGCAATTCAGCAATTC | 1037 | | |
| QY | 882 | CGCTGGGAGTGTGCTGCTTATGGCGGTGGCGACCTTGGCGCATTTGCACTGCTGCTGCTG | 941 | | |
| Db | 1038 | CTGCAAGTGTCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1097 | | |
| QY | 942 | TGTGCGCAACAGATTCGGAGCTGTTTCATGCTGGCGGATTTGCTGGCGGCTTGGCG | 1001 | | |
| Db | 1098 | TGACAAAACAGAGATTCAGACCTCTTTCTTTTGGGTTGCTCACTAGCTGCAAGCGGCTG | 1157 | | |
| QY | 1002 | TCTTTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1061 | | |
| Db | 1158 | TGTTCTTGTGCTATCTATCTGACATACAGGTATATTTGACCAATGAGTTGGAGGT | 1217 | | |
| QY | 1062 | TCTACTGCTGGGATATCTGGCTACGCCAAGATTCACATTTCCCATCATTTGCATCCGCTG | 1121 | | |
| Db | 1218 | TTTATTCATATGAGTATCTGGGTATGCAAAAATACACATTCCAATTTATGATCAGTGT | 1277 | | |
| QY | 1122 | CGGAGCATCAGCCACCATTTGTTCTCTGTTCTTCTTGTGATCTGCACATCTCTGTTGCG | 1181 | | |
| Db | 1278 | CTGAACATCAGCTTACGATGGGTGTTCTTCTTGTGATCTACATATCTTGTATGTA | 1337 | | |
| QY | 1182 | CTTCCAGTGGGAGTGTGTGATCTGATCAAGCAGATCAACGACGAGCGGTTTTCGTGG | 1241 | | |
| Db | 1338 | CTTCCAGCAGGCTTATGGTCTGATCAAAAATATCAACGATGAAGAGATTTTGTGCG | 1397 | | |
| QY | 1242 | TGCTGTACGCTCAGTGGCGGTTTACTTCTGCTGGTGTGGTGGCTTGTGATGTTGATCACC | 1301 | | |
| Db | 1398 | CTCTGTATGCGATCAGTGTGCTGCTTCTTTCGCGAGTGTGTTGCGGCTGATGCTGACTC | 1457 | | |
| QY | 1302 | TCAAGCGGTGGTGTGATCTGCTGGCGGAGTGGCTTTTCGGGACTGTTGGATGTGTTCC | 1361 | | |
| Db | 1458 | TGACCCCGGTGCTGATGCTGCTGGCCATCGCTTCTTCCAATGTTTGTGAGCACTATT | 1517 | | |
| QY | 1362 | TGCAAGAGGATTCGTCTAAGCGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATG | 1421 | | |
| Db | 1518 | TGGGGATGA-----CATGAAAGGGAAACCCACCTGTGGAG-----GACAGCAGTG | 1565 | | |
| QY | 1422 | AAGCTGAGGATTCATTTGAGAGAGAGCGCTGTACGACAGGCTGCGAGCTGGAAGCTGGAAGCATC | 1481 | | |
| Db | 1566 | ATGAGGATGCAAAAGAAACCCAGGAACTTGTATGACAGGCGAGGTAAAGTGAGGAAGC | 1625 | | |
| QY | 1482 | GTACTAAGCATGATGCCAGCAGGATCTGGCGTCTGAGCTCCAACTGGAAGAGTATTGTATA | 1541 | | |
| Db | 1626 | ATGTGACAGAGCAGAGAAACCTGAGAGGCTTGGGCCCAACATCAAAAGCATTTGTGA | 1685 | | |
| QY | 1542 | TTTTGGCGGTTCTAATGCTGTGATGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1601 | | |
| Db | 1686 | CCATGCTGATGCTCATGCTCTGATGATGTTTTCGCGTCCACTGACGCTGGGTGCAAGCA | 1745 | | |
| QY | 1602 | ATGCTACTCTCAGTCCCTCCATTTGCTTGTGCTTTTCCACACAGCTCAAGATGAGTCCCGCA | 1661 | | |
| Db | 1746 | ACGCTACTCTCAGTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1802 | | |
| QY | 1662 | ACATTTAGAGATTTGAGAGGCTTACTTCTGCTTTTCCGAGACACTGCGGATGATG | 1721 | | |
| Db | 1803 | ATATATTAGATGATTTTAGAGAGCGTACTTTTGGTGGAGCAAAACACGGATGAACAGC | 1862 | | |
| QY | 1722 | CTCGGTTATGCTCTTGGTGGGATTTAGGATACAGATAGCGGAAATGGCAAAACAGAACCA | 1781 | | |
| Db | 1863 | CCCGGTCATGTCGTGGTGGGACTACGGCTATCAGATTGCTGGCATGGCCACACAGGACCA | 1922 | | |
| QY | 1782 | CGCTAGTGGATTAATACTGGAACAATAGTCAATAGTGGCTGGTGGTGGTGGTGGTGGTGGTGG | 1841 | | |
| Db | 1923 | CTCTGGTGGATTAACAACCTGGAAACACAGCCACATCGCTGGTGGTGGTGGTGGTGGTGGTGG | 1982 | | |
| QY | 1842 | CTTCAACCGAGGAGAGTCTCTACGAAATATGACATCTCTGACGTTGACTACGTTTGG | 1901 | | |
| Db | 1983 | CTTCCAATGAAAACCGCGCTTAAATCATGAGTCCCTTGTGATGCTGATGTTGG | 2042 | | |
| QY | 1902 | TGATCTTTGGCGGTGATCGGCTATTTCTGCGCATGATATCAACAAGTTTCTCTGTGATGG | 1961 | | |
| Db | 2043 | TTATTTTTCGAGGAGTGTGCTATTTTCGCGGACGATATCAACAAGTTTCTCTGTGATGG | 2102 | | |
| QY | 1962 | TCCGAATTTCTGAGGAGAGCATCCCAAGGACATTAAGGAAGAGGATTAATTTACCGACC | 2021 | | |
| Db | 2103 | TCAGATAGCTGAAGGGGAGCATCCCAAGACATCCGGGAAGTGAATTTTCAACCCAGC | 2162 | | |
| QY | 2022 | CGGTGTAATTCAGGGTAGATGCGAGGTGCTCCGGCCCTGCTCAACTGCTCTTATGTA CA | 2081 | | |
| Db | 2163 | AGGAGAGATTTCCGAGTAGACAAAGCTGGGTCTCTACTCTGTGTTAACTGCTCTTATGTA | 2222 | | |
| QY | 2082 | AATTAAGCTACTACAGATTCGGGGAATTAAGTTTGAATTTGACTACAGAGTCCATCTGGATG | 2141 | | |
| Db | 2223 | AAATGTCTACTACAGATTTGGAGAAATGCAAGCTAGATTTTGCACCTCCCGAGGCTTG | 2282 | | |
| QY | 2142 | ATCGCACAGCTAACCGCGTCAATTTGGGAATGAAGACTTCGATCTGACCTTACCTTGGAGGAGG | 2201 | | |

QY 1302 TCACGCCGCTGCTGTCATCTGCGGAGTGGCTTTTCGGGACTGCTGATGTGTCC 1361
DB 1391 TCACCTCCAGTCTGCTGATGCTGCTGCAATGCTTTTCAATGTTTTGACACTATT 1450
QY 1362 TCACAGAGATTCTGCTTAAGCGAATGGGCAAGCAATAGCGAGCCAGCCAGAGTGATG 1421
DB 1451 TGGGGGATGA-----CATGAAAGGAAATCCACCTGTGGAG-----GACAGCAGTG 1498
QY 1422 AAGCTGAGATTCCATTTCAGAGAGAGACGCTGTAGCAAGAGCTGGCAAGCTCAAGCATC 1481
DB 1499 ATGAGGATGACAAAGAAACCAAGAAATTTGTATGATAGCGCAGGTAAAGTGAAGAAC 1558
QY 1482 GTACTAAGCATGATCCCGAGCAGGATCTGCGGTGAGTCCCACTGAAGAGATTGTTA 1541
DB 1559 ATGCAACTGAACAGGAAAGAACTGAAGAGGATAGGCCCTTAATATATAAAGCAITGTCA 1618
QY 1542 TTTTGGCGCTTCTAATGCTGTTGATGTTGCTGCTCCACTGCACTGGTGGTACACGCA 1601
DB 1619 CCATGTTAGTCTGATGCTATGATGTTGCTGCTCCACTGTACCTGGTCAACGCA 1678
QY 1602 ATGCTACTCCAGTCCCTCCATGTTGCTTGGCTTCCACACAGTCAAGATGATGCCGCA 1661
DB 1679 ATGCTACTCTAGTCCAAAGTGTAGTCTGGCTCATACAA---TCATGATGGCACCAGGA 1735
QY 1662 ACATTTAGCATTTACAGAGGCTTACTACTGCTTTCGCAAGCACTGCCGATGATG 1721
DB 1736 ATATCTTAGATGATTTTAGAGAGCTTACTTTGGCTTAAGCAAAATACAGATGAACATG 1795
QY 1722 CTCGCGTATGCTTTGGTGGATACCGATACAGATAGCGGGAAATGGCAACAGAACGA 1781
DB 1796 CACGAGTATGCTTTGGTGGATATGCTATACAGATAGTGGAAATGGCTAATAGAACTA 1855
QY 1782 CGCTAGTGATAATAATACGTGGCAACATAGTCAATAGGCTGTTGGCAAGCAATGT 1841
DB 1856 CTTGCTGATGATTAACACTGGAATACACCCACATAGACTGCTGGGAAAGCTATGT 1915
QY 1842 CTTCAACGAGGAGAGTCTTACGAAATATGACATCTCTGAGCTGGACTACGTTTGG 1901
DB 1916 CTTCTAATGAACACAGCAGCTTATAAATCATGAGGACTCTAGATGTAGATTATGTTTGG 1975
QY 1902 TCATCTTTGGGCTGATGCTGCTATCTGCGGATGATATCAACAGTCTCTGCGATGG 1961
DB 1976 TTAATTTTGGAGGGTTATGCTATCTGCTGATGATATCAACAAATTTCTCTGATGG 2035
QY 1962 TCCGAATTCTGAGGAGAGATCCCAAGGACATTAAGAAAGCATTAATTTACCGAAC 2021
DB 2036 TTAGGATAGCTGAAGGAGAACATCCCAAGACATTCGGGAAGTCACTATTTTACCCAC 2095
QY 2022 GCGGTGAATTCAGGCTAGATCCGAGAGTGTCCGCGCTGCTCAACTGCTTATGTACA 2081
DB 2096 AGGAGAAATTCGCTGATACAAAGCAGGATCCCTACTTTGTTGATGCTTATGTATA 2155
QY 2082 AATTAAGCTACTACAGATTCGGGGAAATGAATTTGGACTACAGAGTCCATCTGGATATG 2141
DB 2156 AATGCTCATCTACAGATTTGGAGAAATTCGAGTGGATTTTCGTACACCCCAAGTTTG 2215
QY 2142 ATCGCACAGTAAACCGCTCTATCGGATAGGACTTCGATCTGACCTACCTCGAGAGG 2201
DB 2216 ACCGAACAGTAACTGCTGATTTGAAATTAAGACATTAATTCACAACTTTGAAGAG 2275
QY 2202 CTTACACCAAGCACTGCTGTTGTCATCTATAGGGTGAAGAACCCGATGATGTTCA 2261
DB 2276 CTTTACATCAGAACACTGGCTTGTAGGATATATAAGTAAAGACCTGTATACAGGG 2335
QY 2262 ATAGACCATCACTGAAGACCAAGGAGAGACGATCTCCAGCAAACTTCATTTCCAGAA 2321
DB 2336 AGACATTAAGTACCAAACTCGAGTCACCAACATTTTCCCAAAACAGAGTATTTGTCAA 2395
QY 2322 AGAATCTAAGGCTCGCAAG---GGCTACATACGAAACCGACCGGTTGTTGTTAAGGAA 2378
DB 2396 AGAAGACTACCAAAAGGAGGCTGCTACATTAATAAATAAGCTGTTTTTAAGAAAGGCA 2455

QY 2379 AACGAACCTTGAAATAAAC 2397
DB 2456 AGAAATATCTAAGAGAC 2474
RESULT 12
AY074880 2481 bp mRNA linear PRI 10-DEC-2002
LOCUS
DEFINITION Homo sapiens source of immunodominant MHC-associated peptides
(SIMP) mRNA, complete cds.
ACCESSION AY074880
VERSION AY074880.1 GI:19879588
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2481)
AUTHORS McBride K., Baron C., Picard S., Martin S., Boismenu D., Bell A., Bergeron J., and Perreault C.
TITLE The model Bedomi minor histocompatibility antigen is encoded by a mouse homolog of the yeast SRT3 gene
JOURNAL Immunogenetics 54 (8), 562-569 (2002)
MEDLINE 22326278
PUBMED 12439619
REFERENCE 2 (bases 1 to 2481)
AUTHORS McBride K. and Perreault C.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2002) Molecular Biology Group, Compatigene, 6100
Royalmount, Montreal, Qc H4P 2R2, Canada
FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="3"
1. 2481
/gene="SIMP"
1. 2481
/gene="SIMP"
/note="similar to Saccharomyces cerevisiae OSTRY3 and Mus musculus protein encoded by GenBank Accession Number AK018758; source of minor histocompatibility antigen(s)"
/codon_start=1
/product="source of immunodominant MHC-associated peptides"
/protein_id="AAL71884.1"
/db_xref="GI:19879588"
/translation="MAEPSAPESKHSLSNSPWSGLMALGNSRHGHGPGQAQCAHKA
AGGAAPKAPAGLSGCLSGOPAGWQSLLSFTILFLAWLAGFSRRLPAVIRPESIIHEP
DPFNVESTHLASHGFEFLNWFDERAVPLGRIVGTVPLQMTAGLIHWIINTL
NITVHIDVCVFLAPTFSLTSLTRELWNQAGLLAACFTALVPGYISRVAG
SFDNEGTAIFALQFTYLMWKSVTGVSFVMTCCCLSYFYM5AMGWYFIINLPLH
VFVLLMQRXKRYIAYSTFYIVGLISMQIPFVGQFPIRTSEHMAAGVALLQAY
AFQYLDRITKQEFQTLFELGSLAAGAVLSVIYLTGYIAPWSGAFYSLMDTGY
AKHIPTIAGVSEHQPTTWFFDFLILVCTFPAGLWFCIKINIDERVVVALYISA
VYPAGVVRMLTLTPVCMLSAIFSNVFEHVGDMKRENPVDESDSDKNOG
NLXDCKKAKHATQEQTEEGPNIKSTIVTMLMLMLMPFAVHCTVWTSNAYSPPS
VFLASVNDCTRNILDDPEAYFLEQNTDEHARVM5WDYGVQIAGMANRTLYDNN
TWNNSHALVCKSMSSNETAYKIMETLDVDYVLFVFGVIGVGGDDINKFLMWRIA
EGEHPKDIRSDYFTPGEPFRVDKAGSPILLNCIMYKNSYIRFGENWQDFRTPPGFDR
TRNAIGKNKIKFKHLEAFTSEHWLVIYIKAPDNRETLDDHPRVTNIFPKQKYL
KKTTRKRGYIKNKLVEFKGKISKTV"

ORIGIN

Query Match 39, 5%; Score 955.8; DB 9; Length 2481;
Best Local Similarity 64.8%; Pred. No. 1.9e-205;
Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
QY 102 AGGTGGCTGGGTACAGCAGCCTATACCTTCGCCATCTCGCTAATGCTGCTGGCGCG 161
DB 191 AGCGGCTGGGTGGCAGTGGCTTCTCTCTCCATCCATCTCTTCTGGCTGGCTGGCG 250

Db 2396 AGAAGACTACCAAGAGCGTGGCTACATTAATAAATAAGCTGGTTTAAAGAAAGGCA 2455
QY 2379 AACGAACCTTGAATAAATAC 2397
Db 2456 AGAATAATCTAAGAGAGAC 2474

RESULT 13
CQ573675 2855 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 1433 from Patent WO0171042.
DEFINITION CQ573675
ACCESSION CQ573675
VERSION CQ573675.1 GI:41637763
KEYWORDS
SOURCE Drosophila sp.
ORGANISM Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1
AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
TITLE Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 0171042-A 1433 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
1..2855
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN

Query Match 30.3%; Score 733.4; DB 6; Length 2855;
Best Local Similarity 61.0%; Pred. NO. 4.3e-155;
Matches 1302; Conservative 0; Mismatches 761; Indels 72; Gaps 4;

QY 121 CCTATACCTTCGCCATCTCTGCTAATCGCTGCTGGCGGATTTCTCTCGCCTCTT 180
Db 466 CTTGTGAAGCTGGCCATCTCATCTCTGGAGCGGTTTATCATTTGCCACAGCTGTG 525
QY 181 CGCGCTCATCGTTTCGAGTCGATTATCCATGAGTTTATCGGTTTCACTACCGGC 240
Db 526 CTTCTGTGTGCGATTTCGAAGAGCGTAATCCATGAGTTGATCCGTTCACTACCGGC 585
QY 241 CACCGCTACATGTTGCGAGATGTTGTGTAACAATCTCACTGTTTCCAGAGCGGC 300
Db 586 CACCGCGTTTTCGCGAGAGCGGCTTTTCAAGTTCCAACTGTTTCCATGACCGGC 645
QY 301 ATGGTATCGCTTCGCGAGATTTGGCGGTACCGTCTATCCCGGCTGATGATTACGTC 360
Db 646 CTGGTATCCCTTCGCGCGCATCATCGCGGCACCATCTATCCCGGCTGATGCTCACCTC 705
QY 361 CGCGGAATCCATTGGCTGCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTG 420
Db 706 GCGGCGCTGTACCGCTGTATGTTGCTCTCAATGACATTCGACATACGAGCGTGTG 765
QY 421 CGTTTCTTCGCGCGGATCTTCAGTGGCTGACCTTCCATCTCCACCTACCTGCTGACCAA 480
Db 766 CGTCTTCTTCGCGCGCTTCTTCTCTCTGCTGACCAAGCTGGTACCTACGCGCTCACAA 825
QY 481 GGAGCTGTGGTTCGCGGCGCGGCTCTTCGCGCGGCTTCTATCCGATCGCTGCTGG 540
Db 826 GGAGATACACAGACTGGAGCTGGAGCTGGTGGCGCGCTTTGATATCCATCGTTCCCGG 885
QY 541 CTACATCAGTAGTTCGGTGGTGGATCGTACGATTAAGAGGCAATTGCCATATTCGCCCT 600
Db 886 GTATATCTCTCGATCCGTGGCGGATCGTACGACATGAGGCAATCGCCATTTCTGCAAT 945
QY 601 GCAGTTCACCTACTTCTCTGCTGGCTCAGTGAAGCTGGATCCGCTGTTCTGTCGGC 660
Db 946 GCTCTTCCACTACTATTTGTGGATCAAGCGGTAAAGACGGGCGACGATCTTTTGTGCGC 1005
QY 661 CGCAGCGGCTTTGCTCTACTTCTACATGGTGTCCGCTGGGCTGGCTGATGTTTATCAT 720

Db 1006 TATGTCGCGCATTTGGCCCTACTTCTATATGGTCTCTCGTGGGTGGGTATGTCTTCTGAT 1065
QY 721 CAACCTGATACCCCTGCAACGCTTCTGCTACTGCTCATTAATGGGAGGATATCTCGCGGCTCT 780
Db 1066 TAACTTAATCCGCTGCAACGCTGCTGGGCTGATGATCACCGGACGTTTCTCGCACAGAT 1125
QY 781 GCTGACGAGCTACAGACCTTCTACATCTCTGGAGCTGCTGTTCTCCATGAGATCCCTTT 840
Db 1126 CTACATAGCATACAGACGCTATATCTGCTGCGCACCAATCTCTGATGAGATCTCGTT 1185
QY 841 CGTGGGATTCACACCGGATACGACAGTGAACACATGGCTGCGCTGGGAGTGTGTTGCT 900
Db 1186 TGTGGGATTCACACCGGATTCAGAGCTCCGAAACACATGCTGGCACTGGAACTTTGGCCT 1245
QY 901 CTTATGCGCGTGGCCACCTTGGCCATTTGAGTGTGGTCTGCTCGGCAACAGATTCG 960
Db 1246 GTGCGAGATTCACGCTTTCTGCTGATATCTGCGCTCGGCAATTCGCAAGATCACTTGA 1305
QY 961 GAAGCTGTTATCTGCTGGCGGATTTGCTGGGCGTGTGGCTCTTTGTGGCCGCTCGTGT 1020
Db 1306 TCTGCTCTTCAAGACGTTGGTTTCCAGTGTTTTGACTGTGGTGTCTGCTGGGATCCCT 1365
QY 1021 GCTCACATGCTGGCGGTTGTGGCCCGCTGGAGTGAAGCTTCTACTCGCTGGGATPAC 1080
Db 1366 GCTCACGCTTACCGGGAAGTCTCCCGCTGGACCGGAGATTTTACTGCTACTAGATCC 1425
QY 1081 TGGTATGCCCAAGATCCACATTTCCATCATTTGCTGCGAGCATCGGCCACAC 1140
Db 1426 ATCTATGCCAAGATCACATTCATCCATCATCGCTCGGATCGAGCATCAGCTACATC 1485
QY 1141 TTGGTCTCTGTTCTTCTGATCTGACATCTCTGGTGTGGCCCTTCCAGTGGGAGTGTG 1200
Db 1486 CTGTCGCTCTTCTTCTTATTTGATCTGAGATCTCTGCTCTTCTTCTCCGCTGGCTTTA 1545
QY 1201 TACTGTCATCAAGCAGATCAACGACGAGCGGTTTCTGCTGCTGTGAGCCATCAGTGC 1260
Db 1546 TTTCTGCTTTTCCAGCTGACGACTCGACATCTTATTTATTTGATGGCGTACCAG 1605
QY 1261 GGTTTATCTGCTGGTGTGATGGTGTGTTGATGTTGACCCCTCACCGCGGTGTGTGCTAT 1320
Db 1606 TATTTATTTCCCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1665
QY 1321 GCTGCGCGAGTGGCTTTTCCGAGCTGTTGGATGTTGCTCTGCAAGAGGATTCGCTAA 1380
Db 1666 GCTATCCGATTTGCCATTTCCGATCTGCTGGCCAGTATATCAAGAGCTTGA----- 1719
QY 1381 GCGAATGGGCACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCATTTGA 1440
Db 1720 -----TGCCGGAAGCTCATCGAAGCAGGA 1744
QY 1441 GAAAGAGAGCTGTACGCAAGGCTGGCAAGCTGTAAGCTATGATGATGATGATGATGATG 1500
Db 1745 GTAGAGC-----AAGCGCAGCAGCAAGAGCTGGAACA 1779
QY 1501 GCAGATATCTGGCTCAGCTCCAACTGAAAGTATTTGTTATTTTGGCCGCTTCTAATGCT 1560
Db 1780 GCAGACGGGAGCGCTGAGAGCGAGTGGCCATTTGGATTCGTGGGTGTCTATCTCTGAT 1839
QY 1561 GTTGATGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Db 1840 GCTTATAGTGTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1899
QY 1621 CATTTGCTTGGCTTTCCAAACAGTCAAGATGGATCCCGCAACATTTTACAGATTTTACAG 1680
Db 1900 TATTTGTTGAGTGC---CAGSTCGCAGATGGCGCGGCGCATCATTTTTCGATGACTTCCG 1956
QY 1681 AGAGCTTCTACTGCTGCTTTCGAGAACTGCGGATGATGCTGCGGCTTATGCTTGGTG 1740
Db 1957 CGAGGCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2016
QY 1741 GGATTTACGATACAGATAGCGGAATGGCAACAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800

Db 2017 GGACTAGGCTTACCAGATAACGGCCATCGGCAATCGGACGATATTGGTGATAACAATAC 2076
 Qy 1801 GTGGAACAATAGTACATAGCGTGTGGTGGCAAGGCAATGCTCTTCAACCGAGGAGATC 1860
 Db 2077 TTGGAACAACAACATATATCGCGGTGGCGGAGGCGATGGCTCTTTCGGAGGAGAAAC 2136
 Qy 1861 CTACGAATATTGACATCTCTTACGCTGGACTACGTTTGGTGAICTTTGGCGGTGTGAT 1920
 Db 2137 CTACGAGATATAGGGAACCTGATGTGGACTACGTTCTCGTGATTTTCGGAGGCTCAC 2196
 Qy 1921 CGGCTATTCTGGCATGATATCAACAGTTCCTGTGGATGGTCCGAAATGCTGAGG-- 1977
 Db 2197 TGGTACTATCTCGGACATATCAACAAGTTCCTGTGGATGGTCCGATTTGGCGGAGCAC 2256
 Qy 1978 AGAGCATCCCAAGACATTAAGGAAGCGATTACTTTACCGACCGCGTGAATTCAGGGT 2037
 Db 2257 GGATCGTGGTGGCAGATCCGCAAAAGGACTACTATCGGCAACGGAGAGTTCGGAGT 2316
 Qy 2038 AGATGCGGAGGTCTCGGCGCTCGTCACTGCTTATGTAATTAAGCTACTACAG 2097
 Db 2317 GGACAGAGGGCTCACCACTGCTCAATTTGATGTACAGATGTGCTACTATCG 2376
 Qy 2098 ATTGCGGAATTGAAGTTGGACTACAGAGTCCATCTGATATGATCGCACACGTAACGC 2157
 Db 2377 CTTTGGGCAATGTACACGGAAGTGGCAAGCCAGGCGTACGATCGATGTCGCGCG 2436
 Qy 2158 CGTCAATTGGGAATAGGACTTCGATCTGACCTACCTGAGGAGGCTCACACACAGACA 2217
 Db 2437 CGAGATCGGCAACAGAGACTTTGACTGGATGTATCGGAGGAGGCTACACACGAGCA 2496
 Qy 2218 CTGCTTGTGGCTCTATAGGTGGAAGACCGC 2252
 Db 2497 CTGCTGTGGCTCTATACAGTTAAGGATCTGC 2531

RESULT 14

BC046072
 LOCUS 2608 bp mRNA linear : VRT 30-JUN-2004
 DEFINITION Danio rerio integral membrane protein 1, mRNA (cdna clone MGC:56332 IMAGE:5603532), complete cds.
 ACCESSION BC046072
 VERSION 1
 KEYWORDS GIC:28277989
 SOURCE MGC
 ORGANISM Danio rerio (zebrafish)

REFERENCE

1 (bases 1 to 2608)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Suetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.K., Villalón, D.K., Muzny, D., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalski, U., Small, M.A., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 REFERENCE 2 (bases 1 to 2608)
 Strausberg, R.

TITLE

JOURNAL
 PUBMED
 REFERENCE 2 (bases 1 to 2608)
 AUTHORS

TITLE

JOURNAL
 REMARK

COMMENT
 Direct Submission
 Submitted (31-JAN-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Sumio Sugano
 cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 103 Row: 0 Column: 9
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein.

FEATURES

Location/Qualifiers
 1..2608

source
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="MGC:56332 IMAGE:5603532"
 /tissue_type="Whole body, adult male"
 /clone_lib="Sugano SUD adult male"
 /lab_host="DH10B"
 /note="Vector: pME18S-FL3"

gene

1..2608
 /gene="itml"
 /note="Synonyms: cb437, wu:fb38c11, fb38c11"
 /db_xref="LocusID:266797"
 110..2227
 /gene="itml"

CDS

/product="integral membrane protein 1"
 /protein_id="AAH46072.1"
 /db_xref="GI:28277989"
 /db_xref="LocusID:266797"

/translation="MTKLGLRLSEYKQDTLLKLLLSMAAVLSFSTLPSVLRFSV
 IHEFDYFNRTTFLAEEGFKPHNFDRAWYPLGRIITGGITGLMITSALYHV
 LHFPHITDIRNVCVFLAPLSPSTAVTFTKELKADAGALAAAMIAVPGYISR
 SVAGSYDEGIAFCMLITYMKAVKSGSIYSSMCALAYFMVSSMGVYFLNL
 IPLVILMLTGRFSHRIYAVCTVCLGTILSMQISFVGFQPVQSEHMAAFVFL
 COIHAVDYLRSLNAQCFVLRFSVLSVLGILLVGLLMTGKISPMTGFEYIL
 DPSYAKNIPILASVSEHPTWSSYFDLQLLVFPPGLYVCFNNLSDATIFIMY
 GTSYFSAVWRLMLVLPVMSILSGISGVSVLTTPMKLDVSRDKTKKQDSTY
 PKNRSVAGSMILVAFLLITYFTSTWTSSEYSSIVLSARGGDSRIIPDFREA
 YVLRHNTPEDAKVMWMDYQITAMANETILVNNNTWNTHSRVQMASTPEKA
 YELRLDVSYLVIFFGLTGYSDSDINKFLMWVIRIGSTDTGRHKEHDYTPTEGF
 VDRGSPVLLNCLMYKCYRVFGVYTEAKRPPGYDRVARNIGNKDFELDLVBAY
 TTEHLVRIYKVKLDLNRGLST"

ORIGIN

Query Match 29.8%; Score 720.2; DB 5; Length 2608;
 Best Local Similarity 60.8%; Pred. No. 4.1e-152;
 Matches 1299; Conservative 0; Mismatches 743; Indels 93; Gaps 4;
 Qy 118 CAGCTTAATCACTTCGCGCTCTCTAATCGCTGGCGCGATTTTCCTTCGCT 177
 Db 154 CACCTGTGTGAGTGTCTATCTGTCTATGCGACCGTCTCTTCTCCACGACT 213
 Qy 178 CTTGCGCGTCACTTCCTTTCGAGTCGATTCATTCAGTTTCCTGGTTCACATCCG 237
 Db 214 GTTTTCTGTGCTTCGCTTCGAGAGCGTCAACGAGTTTTCATCCGTAATACCG 273
 Qy 238 GGCCACCGCTACATGTTGCGAGATGGTGTGTGTACAACTTCCTCAACTGTTTCGACGAGCG 297

Db 274 GACCACTCGCTCTTGGGAGAGAGGATTTTACAAATTCATACCTGGTTCATATCG 333
QY 298 CGCATGGTATCGGCTCGGACGATTTGGGGGACCGTCTATCCCGGCTGATGATTAC 357
Db 334 GGGTGGTATCCACTTGGGAGGATTTATGGAGGACCATCTATCTGGGTGATGATCAC 393
QY 358 GTCCGGCGGAATCCATTCGGTCTGCGAGTACTCAACATACCGGTCCATATTCGTACAT 417
Db 394 ATCAGAGCTCTATATCAGTCTGCACTTTTCCACATCACCATTTGACATCCGTATGT 453
QY 418 CTGCGTGTCTTGGCGCGGATCTTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC 477
Db 454 CTGTGTGTCTCGCT 513
QY 478 CAAGGAGCTGTGGTCCGGGCGCGGCTCTCTCGCGCGGCTCTCTCTCTCTCTCTCTCT 537
Db 514 CAAGAGCTGAAGGATCGGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 573
QY 538 TGGCTACATCAGTAGGTGGGTGGGATCGTACGATACAGGGGATTCGCCATATTCGC 597
Db 574 TGGATACATCTCCCGTCT 633
QY 598 CCGCAGTTCACCTACTTCT 657
Db 634 CATGCTGTGAGTACTACATGTGGATCAAGGAGTGAAGTCGGGCTCCATTTATTTGGTC 693
QY 658 GCGCGACGCGTCTTCT 717
Db 694 ATCCATGTGTGACCTGGCATATTTTACATGGTCTCTCTCTCTCTCTCTCTCTCTCTCT 753
QY 718 CATCAACTGATACCCCTGACGCTCTGTAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 777
Db 754 GATAAATCTCATCTCTCTGACGCTGTGTGTGATGCTCACCGCGCGTCTCTCTCTCTCT 813
QY 778 TGTGCTGACGCTACAGCACTCTCTACATCTGCGGACCTGCTGTCTCTCTCTCTCTCTCT 837
Db 814 CATCTATGTGGCTACTGACAGTCTACTGCTGGGACCATCTCTCTCTCTCTCTCTCTCT 873
QY 838 CTTCGTGGATTCGAACCGATACGACAGTGAACATGCTGCTGGGAGTCTGCTGGGAGTCTGT 897
Db 874 TTTTGTGATTTCCAGCCCGTCCAGTCTCTGAAACATGCTGCTCTCTCTCTCTCTCTCTCT 933
QY 898 GCTCTTATGCGCGTGGCCACCTTTGGCGCATTTTCAGTTCGGTCTGCTGCGGCAACAGATT 957
Db 934 TCTGTGCCAGATTCACGCTTTTGTGATTAATCTCGGAGCAAACTGAACGCGCAGAGTT 993
QY 958 CCGGAGCTGTTCATCTGCGGAGTTGCTGCTGGGCTGGGCTCTTGTGCGCGTCTGT 1017
Db 994 TGAGGTGTGTTTCAGAGTGTCTATTTCACTGCTGGGATCATCTGCTGCTGCTGCTGCTG 1053
QY 1018 GGTGCTCACCATGCTGGGCTGTGTGGCCCGTGGAGTGGAGCTTCTACTCTGCTGGGA 1077
Db 1054 GTTGTCTATGCTGACAGGAAATCTCTCCCTTGGACTGGCGTTTCTACTCTCTGCTGGA 1113
QY 1078 TACTGGCTACGCAAGATCCATTCCTCATCATTTGCACTCGGTGCGAGGATCAGGCCAC 1137
Db 1114 TCCATGTCAGCCCAAGAACATTCCTCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1173
QY 1138 CACTTGGTCTCTGCTCTCTCTTGTATCTGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
Db 1174 CACATGGTCTCTCTACTTCTGACCTCCAGCTTCTGCTCTCTGCTCTCTGCTCTCTGCTCT 1233
QY 1198 GTGGTACTGCATCAGCAGATCAACGACGAGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTG 1257
Db 1234 TTTACTACTGCTTCAACAACTCTCTGATGCGCAAGATTTATCATCATGATATGGCGTCTAC 1293
QY 1258 TCGGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
Db 1294 CAGCATGTACTTCTCAGCTGTGATGTGGTCTCATGTGCTCTCTCTCTCTCTCTCTCTCT 1353
QY 1318 CATGCTGGCGGAGTGGGCTTTTTCGGGACTGTTGGATGTGTTCTCTGCAAGAGGATTCGTC 1377

Db 1354 CATCTGTCTGGCATTTGGGCTCTC----- 1377
QY 1378 TAAOCGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCAT 1437
Db 1378 ---TCAGTCTCTACCACTTTCATGAAGAACCTGGATGTCAAGTCCAGCCGACAAAGAAAC 1434
QY 1438 TGAGAAGAAGACGCTGTACGACAAGGCTGGCAAGCTGAAGCATCTAAGCATGATGC 1497
Db 1435 CAAGAGCAGCAGAGCTCCACCTACCCATCAAG----- 1468
QY 1498 CCAGCAGGATACTGGCGCTCAGCTCCACCTGAAGATATTGTTATTGTCGCGCTTCTAAT 1557
Db 1469 -----AACGAAGTGGCAAGTGGGATGATTCTAGTGTGCGCTT 1506
QY 1558 GCTGTGTGATGATGTTCTGCTGTCCACTGACGCTGGGTGACCCAGCAATGCTCTACTCCAGTCC 1617
Db 1507 CTTCTCTATCATACATACATTTTCCACTCCACTGGGTGACCCAGGAGGCATATTCGTCGCC 1566
QY 1618 CTCATTTGCTTTGGCTTTCCACAACAGTCAAGATGGATCCCGCAACATTTTAGACGATTT 1677
Db 1567 CTCATCTGCTTTCTGCTC---GCGGTGGCGAGCGGCGCATCATCTTTGATGACTT 1623
QY 1678 CAGAGAGCTTACTACTGCTTTTCGCAAGACACTGCGGATGATGCTCGGCTTATGCTTCTG 1737
Db 1624 CAGAGAGCTTATTTACTGCTCTAGACACACACACAGAGGATGCTTAAGTGTATGCTG 1683
QY 1738 GTGGATTTACGATACAGATAGCGGATGGCAACAGAACAGCAAGCTAGTGGATTAATAA 1797
Db 1684 GTGGGATTTATGCTATCAGATAACAGAGTGGCAATTCGAACGATTTCTGTTGACATAA 1743
QY 1798 TACGTGGAACAATAGTACATAGCGTGGTGGCAAGGCAATGCTTCAACCGAGGAGAA 1857
Db 1744 CACTTGGATTAACACTCACTCTCAGATGCGGTGAGGCAATGGCTCCACAGAGGAGAA 1803
QY 1858 GTCTACGAATTAATGATCTCTTTGACGTGGAACAGTTTGGTGTGATCTTTGCGCGGTGT 1917
Db 1804 ASCCTATGAGATTTATGCGAGAGTGGATGTGAGCTAGCTCTCTGCTGATCTTTGCGCGAT 1863
QY 1918 GATCGCTTATTTCTGGCATGATATCAACAGTTTCTGTGGATGCTCGCAATTCCTGAGGG 1977
Db 1864 GACCGATTTATCTCTGATGATTAATAAGTTCTCTGTGGATGCTGCTGATGCTGCTGCTG 1923
QY 1978 ---AGAGCATCCCAGACATTAAGGAAGCATTAATTTACCGACCGCGGTGAATTCAG 2034
Db 1924 CACAGACACAGGAGGAGCATAAAGAGGACAGATTAATACACCCCACTGGAGATTTCCG 1983
QY 2035 GGTAGATGCGAAGGCTCTCGGCCCTGCTCAACTGCTTATCTATCAAAATTAAGCTACTA 2094
Db 1984 TGTGACCGGAGGCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2043
QY 2095 CAGATTCTGGGAATTAAGTTGGAATACAGAGTCCATCTGGATATGATGCGCACAGTAA 2154
Db 2044 TCGCTTTCGGCCAGGCTTACACCGAAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2103
QY 2155 CGCGCTCATTTGGGAATTAAGCTTTCGATCTGACCTACCTGGAGGAGGCTTACACACAGA 2214
Db 2104 TGCTGAGATTGGCAATTAAGACTTTGAGCTGGAGCTGCTGGAAGAGCTTATACTACAGA 2163
QY 2215 AACTGCTGCTTCTCGCATCTATAGGGTGAAGAAGC 2249
Db 2164 GCACTGGCTGCTCAGGATATACAGGTCAAAGATC 2198

RESULT 15

BC063234

LOCUS

DEFINITION

Danio rerio integral membrane protein 1, mRNA (cDNA clone MGC:77796

IMAGE:7001298), complete cds.

ACCESSION

BC063234

VERSION

BC063234.1

GI:38649116

KEYWORDS

MGC

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

BC063234 2567 bp mRNA linear VRT 30-JUN-2004

Danio rerio integral membrane protein 1, mRNA (cDNA clone MGC:77796

IMAGE:7001298), complete cds.

ACCESSION

BC063234

VERSION

BC063234.1

GI:38649116

KEYWORDS

MGC

SOURCE

Danio rerio (zebrafish)

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 2567)
 Srausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D., Altschul, S.F., Zebard, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Parner, A.A., Rubin, G.W., Hong, L., Scaletton, M., Soares, M.B., Ronald, M.E., Casavant, T.L., Schaefer, A., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

1247932

2 (bases 1 to 2567)
 Srausberg, R.
 Direct Submission
 Submitted (02-DEC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Len Zon, Harvard
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Allison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smalilus, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 147 Row: 1 Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis. Similarity but not identity to protein.

FEATURES

source
 1. 2567
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="MGC:77796 IMAGE:7001298"
 /tissue_type="Whole, adult zebrafish"
 /clone_lib="NIH_ZGC_10"
 /lab_hosts="DH10B"
 /note="vector: pExpress1"
 1. 2567
 /gene="itml"
 /note="synonyms: cb437, wu:fb38c11, fb38c11"
 /db_xref="LocusID:266797"
 70_2187
 /gene="itml"

gene

CDS

/codon_start=1
 /product="Integral membrane protein 1"
 /protein_id="AAH63234.1"
 /db_xref="GI:38649117"
 /translation="MTKGLFLRLSGYKQDITLLKLLLSMAAVLSFSTRFLSVLRPESV
 IHEFPYNYRTTFLAEDEGFKPHNWFDDRAWYPLGRITGGTYPLGLMITSAAAYVH
 LHFHTIDIRNVCVELAPLSPSTALVTYFTKELADAGALLAAATAVVPGYLSR
 SVAGSYDNEGIAJECMLTYMYMKAVKSGSIYSSMCALAYFYVSSRGVYVFLNL
 IFLVLMVLMGRFSRIYVAICTYICGLIILNQISFVGFQVQSEHNAAGVYGL
 COIHAVDYLRSLNAQOFLPFRSIVLSVGLIILSVGLMLTKISLWTRGFRFSL
 DFTYAKNIPITIASVSEHQPTWSSYFDIQLVFPFVGLYYCFNNLSATIFILMY
 GVTSYFSAVNRMLVLAIPVMSIGISGVQLTFPMKLDVSRDKPKTKQKQDSTY
 PIKNEVASCMILVMAFPLITTYFTSTWTSYSSPSIVLSARGGDSRIIFDFFEA
 YWLRNTPEDAKMSWMDYGYQTAMANRTILVDNNTWNTHTISRVQAMASTBEKA
 YEIMRELDVSVLVIFGLTGYSSDDINKFLMWVRIGSDTDGTHIKEHDYTPTEBF
 RVDRGSPVLNCLMYKMCYRFGQVTEAKRPGYDRVNAIGNKDFELDVLEAEY
 TTEHVLVRIYKVKOLDNRGLSRT"

ORIGIN

Query Match 29.7%; Score 718.6; DB 5; Length 2567;
 Best Local Similarity 60.8%; Pred. No. 9.5e-152;
 Matches 1298; Conservative 0; Mismatches 744; Indels 93; Gaps 4;

| | | | |
|----|-----|--|-----|
| QY | 118 | CAGCCTATACCTTCGGCATCTGCTAATCGCTGCGCGGATTTTCCTCGCCT | 177 |
| DB | 114 | CACCTTCGGAAGCTGCTCATCTCTGTATGGAGCGCTGCTGCTCTCCACGACT | 173 |
| QY | 178 | CTTCGCGCTCATCCGTTTCGAGTCGATTATCATGAGTTTTCGCGGTTCACACG | 237 |
| DB | 174 | GTITTCGTGCTTCGGTTCGAGAGCGCTCATCCAGAGTTTTCGATTCGATTCACG | 233 |
| QY | 238 | GGCCACCGCTACAGTGTGAGAAATGTTGTGTAACAATCTCTCACTGTTCCACAGCG | 297 |
| DB | 234 | GACCACTGCTTCCTGGCAGAGAGGGAATTTTACAAATTCATTAATCTGTTGATGTCG | 293 |
| QY | 298 | CGCATGGTATCGCTCGGACGAGATTGTGGCGGTACCGTCTATCCCGCCTGATGATTAC | 357 |
| DB | 294 | GGCTTGGTATCCACTTGGAGGATTATTGGAGGACCATCTATCTCGGTGTTGATGATCAC | 353 |
| QY | 358 | GTCCGGGGGAATCATTGGCTGCTGCGAGTACTCAACATACCGGTCCATATTCGTGACAT | 417 |
| DB | 354 | ATCAGAGCTTATATATACAGTGTGACATCTTCCATCATCACAATGACATCGTATGTT | 413 |
| QY | 418 | CTGCGTGTTCCTGGCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC | 477 |
| DB | 414 | CTGTGTTCCTCGCT | 473 |
| QY | 478 | CAAGGAGCTGTGTCGCGCGCGCGCGCTCTTCGCGCGGAGCTTCATCGCCATCGTCC | 537 |
| DB | 474 | CAAGAGCTGAAGGATGAGCGGCTGGGCTTCTGGCTGCGCTATGATGATGATGATGATG | 533 |
| QY | 538 | TGGTACATCAGTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG | 597 |
| DB | 534 | TGATACATCTCCGTTCTGTGGCAGGCTCTTATGATAATGAAGAAATTCGCAATTTCTG | 593 |
| QY | 598 | CTGCGAGTTCACCTACTTCTCTGCGGTGCGGTCTCAGTGAAGACTGGATCCGCTCTGCTC | 657 |
| DB | 594 | CATGCTGTGAGTACTACATGTGATCAAGGAGTAGTGAAGTCGGGCTCTCATTTATGCTC | 653 |
| QY | 658 | GGCGGAGCGGCTTTGTCCTACTTTCTACATGGTTCGCGCTGGGCTGGCTACGTTGTTAT | 717 |
| DB | 654 | ATCCATGTGTCATTTGGCATATTTTACATGGTTCCTCTCTCTCTCTCTCTCTCTCTCT | 713 |
| QY | 718 | CATCAACCTGATACCCCTGCGAGCTCTCTGCTACTGCTCATATTATGGCAGGTACTCGCGCG | 777 |
| DB | 714 | GATTACCTCATCTCTCTGACGCTGCTGTTGAGTCTACCGCGCGGCTTCTCTCATCG | 773 |
| QY | 778 | TCGTCTGACCAAGTACAGCACCTTTCTACATCTCTGGGACTGCTGTTCTCCATGAGATGCC | 837 |
| DB | 774 | CATCTATGTGGGCTACTGCAGACTCTACTGCTGGGACCATCTCTCTCTCCATGAGATCTC | 833 |
| QY | 838 | CTTCGTGGGATTCACACCGATACGACACCATGACATGCTCGCTGGGAGTGTGTGT | 897 |

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2004, 20:50:43 ; Search time 1108.82 Seconds
(without alignments)
11442.663 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctaaagcgaagaagtgtctg.....ccaaaaaataaaaaaa 2417

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: N Geneseq 23Sep04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002as:*

7: geneseqn2002bs:*

8: geneseqn2003as:*

9: geneseqn2003bs:*

10: geneseqn2003cs:*

11: geneseqn2003ds:*

12: geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|---------------------|
| 1 | 2417 | 100.0 | 2417 | 10 | ADD94789 | ADD94789 Drosophila |
| 2 | 2410.4 | 99.7 | 2699 | 4 | ABL13247 | ABL13247 Drosophila |
| 3 | 2147.4 | 88.8 | 4922 | 4 | ABL13246 | ABL13246 Drosophila |
| 4 | 960.6 | 39.7 | 2710 | 10 | ADD94785 | ADD94785 Mouse SIM |
| 5 | 955.8 | 39.5 | 2481 | 10 | ADD94783 | ADD94783 Human SIM |
| 6 | 733.4 | 30.3 | 2855 | 4 | ABL02795 | ABL02795 Drosophila |
| 7 | 696 | 28.8 | 2953 | 4 | ABL18224 | ABL18224 Drosophila |
| 8 | 635.4 | 26.3 | 2232 | 8 | ABT20789 | ABT20789 Aspergill |
| 9 | 632.8 | 26.2 | 2472 | 10 | ADD94793 | ADD94793 Human ITM |
| 10 | 632.8 | 26.2 | 2472 | 10 | ADH28844 | ADH28844 Human chr |
| 11 | 631.2 | 26.1 | 3094 | 10 | ADD94791 | ADD94791 Mouse ITM |
| 12 | 620.2 | 25.7 | 2466 | 10 | ABE69900 | ABE69900 C. neofor |
| 13 | 615.6 | 25.5 | 2760 | 5 | ABV24502 | ABV24502 Human pro |
| 14 | 562 | 23.3 | 1848 | 8 | ABT18969 | ABT18969 Aspergill |
| 15 | 521.8 | 21.6 | 2733 | 10 | ADD94787 | ADD94787 Yeast STT |
| 16 | 521.6 | 21.6 | 2000 | 10 | ACC61293 | ACC61293 Gene sequ |
| 17 | 521.6 | 21.6 | 2000 | 10 | ADK63659 | ADK63659 Disease t |
| 18 | 519.8 | 21.5 | 2256 | 6 | ABZ32036 | ABZ32036 Candida a |
| 19 | 501.8 | 20.8 | 1464 | 4 | AAK94164 | AAK94164 Human ful |
| 20 | 501.8 | 20.8 | 1864 | 12 | ADL30661 | ADL30661 Full leng |
| 21 | 497 | 20.6 | 1969 | 8 | ABT18375 | ABT18375 Aspergill |

ALIGNMENTS

RESULT 1

ADD94789

ID ADD94789 standard; DNA; 2417 BP.

XX AC ADD94789;

XX XX

XX D7 29-JAN-2004 (first entry)

XX XX

XX Drosophila melanogaster STT3 gene sequence.

XX KW

XX source of immunodominant MHC-associated peptide; SIMP; MHC;

XX major histocompatibility complex; human leukocyte antigen; HLA;

XX cytotoxic; immunosuppressive; antitense therapy; Gene therapy; cancer;

XX lung cancer; intestine cancer; sarcoma; prostate cancer;

XX testicular cancer; breast cancer; melanoma; pancreatic cancer;

XX haematological cancer; immune response; lymphoid cell proliferation;

XX autoimmune disease; transplant rejection; SIMP-derived peptide;

XX fruit fly; gene; ds; STT3.

XX OS Drosophila melanogaster.

XX XX

XX WO2003054008-A2.

XX XX

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX XX

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

XX PI Perreault C, McBride K;

XX DR WPI; 2003-559122/52.

XX DR P-FSDB; ADD94790.

XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic

XX acids and proteins, useful for diagnosing and treating cancer, e.g. lung

XX or breast cancer, or for suppressing an immune response in an autoimmune

XX disease.

XX XX

XX Claim 6; SEQ ID NO 7; 56pp; English.

XX PS This invention relates to a novel isolated or purified human protein,

XX CC

| | | | | | | |
|----|-------|------|------|----|----------|--------------------|
| 22 | 497 | 20.6 | 2603 | 8 | ABT20191 | Abt20191 Aspergill |
| 23 | 497 | 20.6 | 3969 | 8 | ABT17781 | Abt17781 Aspergill |
| 24 | 497 | 20.6 | 4603 | 8 | ABT19595 | Abt19595 Aspergill |
| 25 | 467 | 19.3 | 1209 | 4 | AAH33264 | Aah33264 Human col |
| 26 | 467 | 19.3 | 1209 | 6 | ABL89850 | AbL89850 Human col |
| 27 | 466.4 | 19.3 | 1543 | 2 | AAH85055 | Aah85055 Human sec |
| 28 | 466.4 | 19.3 | 1543 | 2 | ACD18981 | Acd18981 Novel hum |
| 29 | 466.4 | 19.3 | 1543 | 12 | ADG78372 | Adg78372 Human sec |
| 30 | 466.4 | 19.3 | 1543 | 12 | ADN60663 | Adn60663 Human sec |
| 31 | 444.4 | 18.4 | 1371 | 12 | ADP28508 | Adp28508 Human sec |
| 32 | 414.6 | 17.2 | 2510 | 5 | AAF93772 | Aaf93772 Human CDN |
| 33 | 414.6 | 17.2 | 2546 | 2 | AAV44866 | Aav44866 Clone CT5 |
| 34 | 414.6 | 17.2 | 2546 | 5 | AAF98463 | Aaf98463 Human CDN |
| 35 | 408.4 | 16.9 | 2785 | 4 | ABL18208 | AbL18208 Drosophil |
| 36 | 403 | 16.7 | 2537 | 4 | ADH08289 | Adh08289 Human sec |
| 37 | 403 | 16.7 | 2547 | 4 | ADH08315 | Adh08315 Human sec |
| 38 | 403 | 16.7 | 2660 | 6 | ABQ54750 | Abq54750 Human ova |
| 39 | 392.4 | 16.2 | 2284 | 4 | AAH18021 | Aah18021 Human CDN |
| 40 | 373.8 | 15.5 | 2882 | 10 | ADB69539 | AdB69539 C. neofor |
| 41 | 373.8 | 15.5 | 4738 | 10 | ADB69178 | AdB69178 C. neofor |
| 42 | 356.2 | 14.7 | 787 | 4 | AAH07526 | Aah07526 Human CDN |
| 43 | 321.2 | 13.3 | 764 | 5 | AAF93968 | Aaf93968 Primer sp |
| 44 | 310.8 | 12.9 | 1114 | 4 | AAH95794 | Aah95794 Human pro |
| 45 | 279.6 | 11.6 | 507 | 3 | AAC01967 | Aac01967 Human sec |

CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the *Drosophila* melanogaster *Sit3* gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Query Match 100.0%; Score 2417; DB 10; Length 2417;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTCGTCGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
 DB |||||
 QY 1 TCTAAGCGAAGATGTCGTCGATTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
 DB |||||

QY 61 GAATTGGACATGATCGGACCGCGAAGATGCTGAACAGCAGGTCGGCTACAGCAG 120
 DB |||||

QY 61 GAATTGGACATGATCGGACCGCGAAGATGCTGAACAGCAGGTCGGCTACAGCAG 120
 DB |||||

QY 121 CCTAATACCTTCGCCATCTCTGTAATCGCTCGCTGCGCGGATTTCTCGCTCTT 180
 DB |||||

QY 121 CCTAATACCTTCGCCATCTCTGTAATCGCTCGCTGCGCGGATTTCTCTCTGCTCTT 180
 DB |||||

QY 181 GCGGTATCCGTTTCGAGTCGATATCCATGAGTTTGATCCGTTCACTACCGGCTC 240
 DB |||||

QY 181 GCGGTATCCGTTTCGAGTCGATATCCATGAGTTTGATCCGTTCACTACCGGCTC 240
 DB |||||

QY 241 CACGCGTACATGTCGAGATGTCGTACACTCCCTCACTGTCGTCGAGCGGCGC 300
 DB |||||

QY 241 CACGCGTACATGTCGAGATGTCGTACACTCCCTCACTGTCGTCGAGCGGCGC 300
 DB |||||

QY 301 ATGGTATCCGTCGCGAGATGTCGCGGTAACGTCATCCGCGCTGATGATACGTC 360
 DB |||||

QY 301 ATGGTATCCGTCGCGAGATGTCGCGGTAACGTCATCCGCGCTGATGATACGTC 360
 DB |||||

QY 361 CGGCGGAATCCATTCGCTGTCGACGTCATCAACATACCGGTCATATTCGTCATCTG 420
 DB |||||

QY 361 CGGCGGAATCCATTCGCTGTCGACGTCATCAACATACCGGTCATATTCGTCATCTG 420
 DB |||||

QY 421 CGTGTCTCGCGCGGATTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480
 DB |||||

QY 421 CGTGTCTCGCGCGGATTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 480
 DB |||||

QY 481 GAGCTGTGTCGCGCGGCGGCGCTTCGCGCGGCGGCGCTTCGCGCGGCGGCGCTTCG 540
 DB |||||

QY 481 GAGCTGTGTCGCGCGGCGGCGCTTCGCGCGGCGGCGCTTCGCGCGGCGGCGCTTCG 540
 DB |||||

QY 541 CTATCATAGTAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 600
 DB |||||

QY 541 CTATCATAGTAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 600
 DB |||||

QY 601 GCAGTTTACCTTCT 660
 DB |||||

QY 601 GCAGTTTACCTTCT 660
 DB |||||

QY 661 CGCAGCGCTTCT 720
 DB |||||

QY 661 CGCAGCGCTTCT 720
 DB |||||

QY 721 CAACCTGATACCCCTGACGTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 780
 DB |||||

DB |||||

QY 721 CAACCTGATACCCCTGACGTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 780
 DB |||||

QY 781 GCTGACCACTACAGCACTTCTACATCCTGGGACTGCTGTCCTCCATGAGATCCCTCTT 840
 DB |||||

QY 781 GCTGACCACTACAGCACTTCTACATCCTGGGACTGCTGTCCTCCATGAGATCCCTCTT 840
 DB |||||

QY 841 CPTGGGATTCACCACTACAGCACTTCTACATCCTGGGACTGCTGTCCTCCATGAGATCCCTCTT 900
 DB |||||

QY 841 CPTGGGATTCACCACTACAGCACTTCTACATCCTGGGACTGCTGTCCTCCATGAGATCCCTCTT 900
 DB |||||

QY 901 CCTTATGCGCTGGGCGCTTCTGCGCATTTGAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 960
 DB |||||

QY 901 CCTTATGCGCTGGGCGCTTCTGCGCATTTGAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 960
 DB |||||

QY 961 GAAGCTGTTTCACTGTCGCGGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1020
 DB |||||

QY 961 GAAGCTGTTTCACTGTCGCGGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1020
 DB |||||

QY 1021 GCTCACCATGCTGGGCGCTTCTGCGCATTTGAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1080
 DB |||||

QY 1021 GCTCACCATGCTGGGCGCTTCTGCGCATTTGAGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1080
 DB |||||

QY 1081 TGCTACGCAAGATCCACATTTCCCATTCATTCGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1140
 DB |||||

QY 1081 TGCTACGCAAGATCCACATTTCCCATTCATTCGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1140
 DB |||||

QY 1141 TTGGTTCTCGTTCTTCTTTCGATTCGCACTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1200
 DB |||||

QY 1141 TTGGTTCTCGTTCTTCTTTCGATTCGCACTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1200
 DB |||||

QY 1201 GTACTGCATCAAGCAGATCAAGCAGCAGCGGCTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1260
 DB |||||

QY 1201 GTACTGCATCAAGCAGATCAAGCAGCAGCGGCTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1260
 DB |||||

QY 1261 GGTTCATTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1320
 DB |||||

QY 1261 GGTTCATTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1320
 DB |||||

QY 1321 GCTGGCGGAGTGGCTTTTCGCGACTGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1380
 DB |||||

QY 1321 GCTGGCGGAGTGGCTTTTCGCGACTGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1380
 DB |||||

QY 1381 GCGAATGGGCAAGCCATTAAGCGGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
 DB |||||

QY 1381 GCGAATGGGCAAGCCATTAAGCGGAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1440
 DB |||||

QY 1441 GAAGAAGACGCTGTACGCAAGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCCCA 1500
 DB |||||

QY 1441 GAAGAAGACGCTGTACGCAAGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCCCA 1500
 DB |||||

QY 1501 GCAGGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1560
 DB |||||

QY 1501 GCAGGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1560
 DB |||||

QY 1561 GTTGATGATGTTTCGCTGTCCTCACTGCACTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1620
 DB |||||

QY 1561 GTTGATGATGTTTCGCTGTCCTCACTGCACTGGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1620
 DB |||||

QY 1621 CATTCG 1680
 DB |||||

QY 1621 CATTCG 1680
 DB |||||

QY 1681 AGAGGCTTACTACTGCTGCTTCGAGAACACTGCGCATGATGCTCGCGTTAATGCTGCTGCTGCTG 1740
 DB |||||

QY 1681 AGAGGCTTACTACTGCTGCTTCGAGAACACTGCGCATGATGCTCGCGTTAATGCTGCTGCTGCTG 1740
 DB |||||

QY 1741 GGNATACGATACAGATAGCGGAAATGGGAAACAGACAGCAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1800
 DB |||||

QY 1741 GGNATACGATACAGATAGCGGAAATGGGAAACAGACAGCAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1800
 DB |||||

QY 1801 GTGGAAACAATAGTCAATAGCGCTGGTGGCAAGCAATGCTTCAACCGGAGGAGATC 1860
 DB |||||

Db 730 CAACCTGATACCCCTGACAGCTCTTCGTAAGCTCATATGGGAGGACTCTCGCCGCTCT 789
Qy 781 GCTGACCAAGCTACAGACCTTCTACATCTCGGAGCTGCTGTTCTCCATGACAGATCCCTT 840
Db 790 GCTGACCAAGCTACAGACCTTCTACATCTCGGAGCTGCTGTTCTCCATGACAGATCCCTT 849
Qy 841 GCTGAGATCCAAACGATACGCAACAGTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 850 GCTGAGATCCAAACGATACGCAACAGTGAACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 909
Qy 901 CTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 910 CTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 969
Qy 961 GAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 970 GAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1029
Qy 1021 GCTACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1030 GCTACCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1089
Qy 1081 TGCTACGCAAGATTCACATTCCTCATTCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1090 TGCTACGCAAGATTCACATTCCTCATTCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1149
Qy 1141 TTGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1150 TTGGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1209
Qy 1201 GTACTGCAAGCAAGATTCACATTCCTCATTCGATTCGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 1210 GTACTGCAAGCAAGATTCACATTCCTCATTCGATTCGCTGCTGCTGCTGCTGCTGCTGCT 1269
Qy 1261 GCTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1270 GCTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1329
Qy 1321 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
Db 1330 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1389
Qy 1381 GCGAATGGCAAGCAAGATTCACATTCCTCATTCGATTCGCTGCTGCTGCTGCTGCTGCT 1440
Db 1390 GCGAATGGCAAGCAAGATTCACATTCCTCATTCGATTCGCTGCTGCTGCTGCTGCTGCT 1449
Qy 1441 GAAGAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Db 1450 GAAGAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1509
Qy 1501 GCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Db 1510 GCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1569
Qy 1561 GTTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1570 GTTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1629
Qy 1621 CATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Db 1630 CATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1689
Qy 1681 AGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 1690 AGAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1749
Qy 1741 GGATTAACGATACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Db 1750 GGATTAACGATACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1809
Qy 1801 GTGGACAAATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860

Db 1810 GTGGACAAATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1869
Qy 1861 CTACGAAATTTATGACATCTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1870 CTACGAAATTTATGACATCTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1929
Qy 1921 CGCTTATTTCTGGGATGATATCAACAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Db 1930 CGCTTATTTCTGGGATGATATCAACAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1989
Qy 1981 GCATCCAAAGGACATTAAGGAAAGCGATTTACTTTACCGACCGGCTGCTGCTGCTGCTGCT 2040
Db 1990 GCATCCAAAGGACATTAAGGAAAGCGATTTACTTTACCGACCGGCTGCTGCTGCTGCTGCT 2049
Qy 2041 TGCCGAAGTGTCTCGGCCCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
Db 2050 TGCCGAAGTGTCTCGGCCCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2109
Qy 2101 CGGGGAATTTGAAGTTGGACTACAGAGGTCCTATCTGGATATGATCGCACACGTAACGCCGT 2160
Db 2110 CGGGGAATTTGAAGTTGGACTACAGAGGTCCTATCTGGATATGATCGCACACGTAACGCCGT 2169
Qy 2161 CATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
Db 2170 CATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2229
Qy 2221 GCTTGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
Db 2230 GCTTGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2289
Qy 2281 CAAGGAGACAGATTCCTTCCAGCAAACTTCATTTCCGAGGAGGCTGCTGCTGCTGCTGCTGCT 2340
Db 2290 CAAGGAGACAGATTCCTTCCAGCAAACTTCATTTCCGAGGAGGCTGCTGCTGCTGCTGCTGCT 2349
Qy 2341 GGCTACATACGAAACCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
Db 2350 GGCTACATACGAAACCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2409
Qy 2401 AAAAAAATAA 2412
Db 2410 AAAAAAATAA 2421

RESULT 3
ABL13246/s
ID ABL13246 standard; cDNA; 4922 BP.
XX ABL13246;
XX AC ABL13246;
XX DT 26-MAR-2002 (first entry)
XX DB Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-FSDS; ABB69143.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX Claim 1: SEQ ID NO 34220; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIFO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;

Query Match 88.8%; Score 2147.4; DB 4; Length 4922;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 1; Indels 223; Gaps 4;

1 TCTAAGCGAAGATGTGCTGCTATTCAGATCGTTATATATTTTCGAGTTACTGCTG 60
3913 TCTAAGCGAAGATGTGCTGCTATTCAGATCGTTATATATTTTCGAGTTACTGCTG 3854
61 GAATGGGACATGAATCGGACGCCGAGATGCTGAACAGCAAGTGTGCTGCTACACAG 120
3853 GAATGGGACATGAATCGGACGCCGAGATGCTGAACAGCAAGTGTGCTGCTACACAG 3794
121 CTAATACCTTCGCTATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTGCGCTCT 180
3793 CTAATACCTTCGCTATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTGCGCTCT 3734
181 CGCGTCACTCGTTTCGAGTCGATTCATCGATGATTCGATCGGTTCAATACCGGCG 240
3733 CGCGTCACTCGTTTCGAGTCGATTCATCGATGATTCGATCGGTTCAATACCGGCG 3674
241 CACCGCTACATGTTGCGAAGTGTGTTGATCACTTCCTCACTGTTTCGACGAGCGCG 300
3673 CACCGCTACATGTTGCGAAGTGTGTTGATCACTTCCTCACTGTTTCGACGAGCGCG 3614
301 ATGGTATCGCTCGGAGGATTTGGCGGTACCGTCTATCCCGGCTGATGATACGTC 360
3613 ATGGTATCGCTCGGAGGATTTGGCGGTACCGTCTATCCCGGCTGATGATACGTC 3554
361 CGGCGAATCCATTTGGCTGCTCAGTACTCAACATCCGTTCCATATTCGTGACATCTG 420
3553 CGGCGAATCCATTTGGCTGCTCAGTACTCAACATCCGTTCCATATTCGTGACATCTG 3494
421 CGTGTTCCTGGCGCGATTTTCAGTGGCTGACCTTCATCTCCACCTACCTGTGACCAA 480
3493 CGTGTTCCTGGCGCGATTTTCAGTGGCTGACCTTCATCTCCACCTACCTGTGACCAA 3434
481 GGAGCTGTGTTCCGCGGCGCGGCTCTTGGCGGCGAGTTTCATCGCATCTGTCGCTGG 540
3433 GGAGCTGTGTTCCGCGGCGCGGCTCTTGGCGGCGAGTTTCATCGCATCTGTCGCTGG 3374
541 CTACATCAGTAGTCTGCTGCTGATTCGTAAGAGAGGCTATTCGCAATTCGCGCT 600
3373 CTACATCAGTAGTCTGCTGCTGATTCGTAAGAGAGGCTATTCGCAATTCGCGCT 3314
601 GCAGTTCACTTCTCTGCTGGTGGCTTCAGTGAAGACTCGATCCGTTCTGTCGGC 660
3313 GCAGTTCACTTCTCTGCTGGTGGCTTCAGTGAAGACTCGATCCGTTCTGTCGGC 3254
661 CGCAGCGCTTTGCTCTACTTCTAATGTTGCGCTGGGTTGGCTACGTTTCATCAT 720
3253 CGCAGCGCTTTGCTCTACTTCTAATGTTGCGCTGGGTTGGCTACGTTTCATCAT 3194
721 CAACCTGATACCCCTGCGCTCTTCGTACTGCTCATTTATGGCAGGTACTCGCGCGCT 780

Db 3193 CAACTGATACCCCTGACGCTCTTCGATCTGCTCATTTATGGGAGGTACTCGCGGCTCT 3134
QY 781 GCTGACAGCTACAGCACCTTCTACATCTCTGGAGTCTCTTCTCCATGAGATCCCTT 840
Db 3133 GCTGACAGCTACAGCACCTTCTACATCTCTGGAGTCTCTTCTCCATGAGATCCCTT 3074
QY 841 GCTGGGATTCACACCGGATAGCACCATGAGTCTGCTGCTGCTGCTGCTGCTGCT 900
Db 3073 GCTGGGATTCACACCGGATAGCACCATGAGTCTGCTGCTGCTGCTGCTGCTGCT 3014
QY 901 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 3013 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2954
QY 961 GAAGCTGTTTCATCTGCTGCGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 2953 GAAGCTGTTTCATCTGCTGCGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2894
QY 1021 GCTCAACATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 2893 GCTCAACATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2834
QY 1081 TGGCTACGCCAAGATCCACATTTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 2833 TGGCTACGCCAAGATCCACATTTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2774
QY 1141 TTGCTTCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 2773 TTGCTTCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2714
QY 1201 GTACTGCTCAAGCAGATCAACGAGCGGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
Db 2713 GTACTGCTCAAGCAGATCAACGAGCGGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 2654
QY 1261 GGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 2653 GGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2594
QY 1321 GCTGCGCGAGTGGCTTTTCTGCGACTGTTGGATGTTTCTCTCAAGAGGATTCGCTAA 1380
Db 2593 GCTGCGCGAGTGGCTTTTCTGCGACTGTTGGATGTTTCTCTCAAGAGGATTCGCTAA 2534
QY 1381 GCGAATGGGACAGCCATTAAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1440
Db 2533 GCGAATGGGACAGCCATTAAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 2474
QY 1441 GAAAGAGCGCTGTACGACA----- 1460
Db 2473 GAAAGAGCGCTGTACGACAAGGTGAGTCTTACTAACACACATCCATGGGATTTGATTT 2414
QY 1461 -----AGGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 1501
Db 2413 TAAATATACATCGCATTTTGGAGCTGGCAAGCTGAAGCATCGTACTAAGCATGATGCCAG 2354
QY 1502 CAGGATCTGCGCTCAGCTCCCACTCAAGAGTATTTGTTATTTTGGCGCTTTCTAATGCTG 1561
Db 2353 CAGGATCTGCGCTCAGCTCCCACTCAAGAGTATTTGTTATTTTGGCGCTTTCTAATGCTG 2294
QY 1562 TTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1621
Db 2293 TTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2234
QY 1622 ATTGCTTGGCTTTTCCCAACAGTCAAGATG----- 1652
Db 2233 ATTGCTTGGCTTTTCCCAACAGTCAAGATGTTGGTAAAGCAACAGTAGCATCATGCTTCT 2174
QY 1653 -----GATCCCGCAACATTTTAGACGATTTTACAGAG 1684
Db 2173 TTAACCTATCTAACCTTTTCTCCACTAGATCCCGCAACATTTTAGACGATTTTACAGAG 2114
QY 1685 GCTTACTACTGCTTTTCGAGAACACTCGCGATGATGCTCGCTGTTATGCTGCTGGGAT 1744

Db 2113 GCTTACTACTGGCTTCCGAGAACACTGCGGATGATGCTCGCGTTAATGCTTGTGGTGGAT 2054
 Qy 1745 TACGGATACAGATAGCGGGAATGGCAAAACAGAACGACGCTAGTGGATTAATAATACGTGG 1804
 Db 2053 TACGGATACAGATAGCGGGAATGGCAAAACAGAACGACGCTAGTGGATTAATAATACGTGG 1994
 Qy 1805 AACATAGTACATAGCGTGGTGGGAGGCAATGCTTCAACGAGGAGAGTCTTAC 1864
 Db 1993 AACATAGTACATAGCGTGGTGGGAGGCAATGCTTCAACGAGGAGAGTCTTAC 1934
 Qy 1865 GAAATATGACATCTCTTGAAGTGGGAGTCTTGTGATCTTGTGGCGTGTGATCGGC 1924
 Db 1933 GAAATATGACATCTCTTGAAGTGGGAGTCTTGTGATCTTGTGGCGTGTGATCGGC 1874
 Qy 1925 TATTCGCGGATGATACAAAGTCTCTGATGCTCGGATGCTCGGATGCTCGGAGGAGAT 1984
 Db 1873 TATTCGCGGATGATACAAAGTCTCTGATGCTCGGATGCTCGGATGCTCGGAGGAGAT 1814
 Qy 1985 CCCAAGGACATTAAGGAAGGAGTCTTACCGACGCGGTGAATTCAGGGTAGATGCC 2044
 Db 1813 CCCAAGGACATTAAGGAAGGAGTCTTACCGACGCGGTGAATTCAGGGTAGATGCC 1754
 Qy 2045 GAAGTGTCTCGGCGCTCTCAAGTCTTATGTACAAATTAAGTCTACTACAGATTCGGG 2104
 Db 1753 GAAGTGTCTCGGCGCTCTCAAGTCTTATGTACAAATTAAGTCTACTACAGATTCGGG 1694
 Qy 2105 GAATTAAGTGGACTA----- 2121
 Db 1693 GAATTAAGTGGACTACAGTAAGCGGAACATTTCTCAGGTAGCGATGCACACTAA 1634
 Qy 2122 -----CAGAGTCCATCTCGATATGATCGCACAGTACGCGCTCANTGGGATA 2171
 Db 1633 CTGCTCTTTCAGAGTCCATCTCGATATGATCGCACAGTACGCGCTCANTGGGATA 1574
 Qy 2172 AGSACTTCGATCTGACCTACCTGAGGAGGCTACACACAGAACACTGCTTGTTCGCA 2231
 Db 1573 AGSACTTCGATCTGACCTACCTGAGGAGGCTACACACAGAACACTGCTTGTTCGCA 1514
 Qy 2232 TCTATAGGTGAAGAGCGCATGATGATCAATAGACCATCATCAGAGACCAAGGAGAGAA 2291
 Db 1513 TCTATAGGTGAAGAGCGCATGATGATCAATAGACCATCATCAGAGACCAAGGAGAGAA 1454
 Qy 2292 CGATTCCTCCAGCAAACTTCAATTCGAGAA----- 2321
 Db 1453 CGATTCCTCCAGCAAACTTCAATTCGAGAAAGGTAGTGTACCCTCGAGCTCCCAATGAC 1394
 Qy 2322 -----AGACTCTAAGCGTCCGAGGCTCATACGAAACC 2357
 Db 1393 TTCATTAACTCTGATTTCTTGGAGAACTCTAAGCGTCCGAGGCTCATACGAAACC 1334
 Qy 2358 GACCGGTTGTTTGAAGGAGAAACGAACTTGAATAAACCACCAAAAAAATAA 2412
 Db 1333 GACCGGTTGTTTGAAGGAGAAACGAACTTGAATAAACCACCAAAAAAATAA 1279

RESULT 4
 ADD94785

ID ADD94785 standard; cDNA; 2710 BP.

XX AC ADD94785;

XX AC ADD94785;

XX DT 29-JAN-2004 (first entry)

XX DE Mouse SIMP cDNA sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cystostatic; immunosuppressive; antitense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ss.

XX OS Mus musculus.
 XX FN W02003054008-A2.
 XX PD 03-JUL-2003.
 XX PF 18-DEC-2002; 2002WO-CA001967.
 XX PR 20-DEC-2001; 2001US-00028394.
 XX PA (COMP-) COMPATIGENE INC.
 XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX P-PSDB; ADD94785;

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

XX Disclosure; SEQ ID NO 3; 69pp; English.

XX This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcoma, prostate cancer, testicular cancer, breast cancer, melanoma, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the mouse SIMP cDNA which is related to the invention. Note: this sequence does not appear in the specification but was obtained by the indexer from GenBank.

XX SQ Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

Query Match 39.7%; Score 960.6; DB 10; Length 2710;
 Best Local Similarity 64.8%; Pred. No. 8, 2e-244;
 Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4;

Qy 102 AGTGGCTGGCTACAGAGCCCTAAATCACCTTGGCCATCTCTAATCGCTGGCTGGCCG 161
 Db 258 AGCGGCGGGTGGCAGTGGTCTCTCTCCATCTCTCTCTGGCTGGCTGGCTGGCCG 317
 Qy 162 GATTTCT 221
 Db 318 GCTTCAGCTCGGCGCT 377
 Qy 222 CGTGGTTCAACTACCGGCGCACCGGCTACATGTCGAGAAATGGTTGGTACAACTCTCTCA 281
 Db 378 CGTGGTTCAACTACATAGATCAACACATCATCTCTGATCTCTATGAGTTCTTAA 437
 Qy 282 ACTGGTTTCAGACGAGCGCATGATCTCGCTCGGAGGATTTGGGCGGTACCGCTCTATC 341
 Db 438 ATTGGTTTGAAGAGAGCATGTTACCCACTGGGAGAGATAGTGGTGGACCGCTTACC 497
 Qy 342 CCGGCTCTGATGATAGTCTCGGCGGAATCCATTTGCTCTCTCTCTCTCTCTCTCTCTCT 401
 Db 498 CAGGCTTGTATGATTAACAGCTGGCTTATTCATTTGATTTAAATACATTTGAACATAACAG 557
 Qy 402 TCCATATTCTGACATCTCGCTGTTTCTTGGGCGCCATCTCTCAGTGGCTGACCTCTCTCT 461
 Db 558 TTCATATAGAGATGTGTGTATTCCTTTGACACCAACTTTTAGCGGCTTACATCATAT 617

QY 462 CCACCTACCTGCTGACCAAGAGCTGTGGTCCGGGGGCGCGGCTCTTCCGGCGGACGCT 521
Db 618 CTAGCTTCCTGCTAAGAGAGCTGTGGAACCAAGAGAGAGAGCTTCTAGCTGCTGCT 677
QY 522 TCATCGCCATCGTCTGCTGCTACATCAGTAGGTGGTGGTGGATCGTAGAAGAGG 581
Db 678 TCATCGCTATCATCAGAGGTACATATCTCGGTGAGTGGCGGATCCCTTGTAAATGAAG 737
QY 582 GCATTGCCATATTCGCCCTGCAAGTTCACCTACTTCTCTGCGGTGCGGTGAGTGAAGACTG 641
Db 738 GCATTGCCATATTTGCGCTTCAGTTTCACTTACTACTTATGGGTAAAGTCTGTGAAGACG 797
QY 642 GATCGGTGCTGCTGGCGGCGGAGCGGCTTGTCTACTTCTATCATGGTGTCCGCTGG 701
Db 798 GGTCTGTGCTTGGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857
QY 702 GTGGCTACGTGCTATCATCAACCTGATACCCCTGCGAGCTCTTGGTACTGCTCATATGG 761
Db 858 GAGGTATGCTTATCATCAACCTCATCCCTCTCCATGCTGTTGTTGCTGCTGATGC 917
QY 762 GCAGGTACTCGCGGCTGCTGACAGCTACAGACCTTCTATCATCTGGGACTGCTGT 821
Db 918 AGAGGTACAGCAAGAGAGTATCATAGCATATAGCATTGTGATCATTTGTTGGGTTTAAAT 977
QY 822 TCTCATGCGATCCCTTCTGTTGGATTCACACCGGATACGACAGTGAACACATGGCTG 881
Db 978 TATCATGCGATACCTTTGTTGGATTTTCAGCCAAATCAGAAACAGCGAGCACATGGCAG 1037
QY 882 CGCTGGAGTGTGCTGCTTATGCGGTGGCCACCTTGGCCATTTGCACTGCTGCTGCTG 941
Db 1038 CTGAGGTGCTTCTGCGCTGCTGCAAGCTTACGCTTTTTCAGATATCTGAGAGACGGT 1097
QY 942 TGTGCGCAACAGATTCGGAGCTGTTTCATCGTGGCGGATTCGCTGGTGGGCTTGGCG 1001
Db 1098 TGACAAAACAGAGTTCAGACCTTCTTCTTTTGGGTGCTCTACTAGCTGCGAGGCGCTG 1157
QY 1002 TCTTTGTGCGCTGCTGCTGCTACCATGCTGGGCTGTGGCCGCTGGAGTGGAGCT 1061
Db 1158 TGTCTCTTAGTGTCTATCTGACATACAGAGTATATTCACCATGAGTGGAGTGGAGT 1217
QY 1062 TCTACTGCTGTTGGATCTGCTACGCAAGATCCCATTCCTCATTTGCACTGCTGCTG 1121
Db 1218 TTTATTCATCTAGGATCTGGTATGCAAAATACACATTTCCAAATTTTTCATCAGTGT 1277
QY 1122 CGAGCATCAGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1181
Db 1278 CTGAACATCAGCTTACGATGGGTGCTTCTTCTTGTATCTACATTTCTTGTATGTA 1337
QY 1182 CCTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241
Db 1338 CCTTCCAGCGGCTATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1397
QY 1242 TGCTGTACCCATGCTGGGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
Db 1398 CTCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1457
QY 1302 TCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361
Db 1458 TGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1517
QY 1362 TGCAGAGGATTCGCTTAAAGGATGGGCAAGCATAGCGGACCCAGGAGTGGATG 1421
Db 1518 TGGGGATGA-----CATGAAAGGGGAAACCCACCTGCTGGAG-----GACAGCAGT 1565
QY 1422 AAGCTGAGGATTCATTTAGAGAGAGAGCTGTACGACAAAGCTGGCAAGCTGAAGCATC 1481
Db 1566 ATGAGGATGACAAAAGAAACCCAGGAACTTGTATGACAAAGCAGGTAAAGTGAAGAGC 1625
QY 1482 GTACTAAGCATGATGCCAGAGGATCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 1541
Db 1626 ATGTGACAGCAAGAGAAACCTGAAGAGGCTTGGGCCCCCAACATCAAAAGCATTTGTA 1685

QY 1542 TTTTGGCGGTTCTAATGCTGTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1601
Db 1686 CCACTGCTGATGCTCATGCTCTCTGATGATGTTCCGGTCCACTGCACTGCTGCTGCTGCT 1745
QY 1602 ATGCTACTCTCAGTCCCTCCATTTGCTTGGCTTTCCACACAGCTCAAGATGGATCCGCA 1661
Db 1746 AGCCTACTCTCAGTCCAGTGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1802
QY 1662 ACATTTAGACGATTTTCAGAGAGGCTTACTACTGCTTTCGAGAACACTGCGGATGATG 1721
Db 1803 ATATATTAGATGATTTTAGAGAGCGTACTTTTGGCTGAGACAAACACGGATGAACAG 1862
QY 1722 CTCGGTATCTCTGTTGGGATTTACGATACAGATAGCGGAAATGGCAACACAGACCA 1781
Db 1863 CCCGGTCTATGCTGCTGGGCTTACGGCTATCAGATTTGCTGCTGCTGCTGCTGCTGCTG 1922
QY 1782 CGCTAGTGATTAATAATACCTGGAACAAATAGTCACTAGCGCTGGTGGCAAGCAATGT 1841
Db 1923 CTCTGGTGAATAACACACCTGGAAACACAGCCACATCGCACTGGTTCGGAAGCTATGT 1982
QY 1842 CTTCAACCGAGGAGAACTCTTACGAATATGATCTCTTACGCTGGGACTAGCTTTTGG 1901
Db 1983 CTTCAATGAACCGCGCTTATATAATCATAGGCTCCCTTGTGCTGCTGCTGCTGCTGCTG 2042
QY 1902 TGATCTTTGGCGGTGTGATCGGCTATTCTCGCGATGATFCAACAAAGTTCCTGTGGATG 1961
Db 2043 TTAATTTTCGAGGAGTGAATGGCTATTTCGGGGACGATATCAACAAAGTTCCTCTGGATG 2102
QY 1962 TCCGAATGCTGAGGAGAGCCTCCCAAGGACATTAAGGAACGATTAATTTTACCAGC 2021
Db 2103 TCAGATAGCTGAGAGGAGCCTCCCAAGACATTCGGGAGTGAATTTTCCACCAGC 2162
QY 2022 CGGTGAATTCAGGTTAGATGCGAAGTCTCCGGCCCTGCTCAACTGCTTATCTACA 2081
Db 2163 AGGAGAGTTCGAGTGAACAAAGCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2222
QY 2082 AATTAAGCTACTACAGATTCGGGGAATGAAGTTGGACTACAGAGTCCCATCTGGATG 2141
Db 2233 AATGTCTACTACAGATTTGGGAATGCACTGATTTTCGCACTCCCCAGGCTTTG 2282
QY 2142 ATCGCACACGPAACCGCTCATTTGGGAATGAAGCTTCGATCTGCACTACCTGGAGAG 2201
Db 2283 ACCGAACACGTAATGCTGAGATTTGGAATTAAGACATTAATTTCAAGCATTTTGGAGAG 2342
QY 2202 CCTACACACAGAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2261
Db 2343 CTTTATCATCAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2402
QY 2262 ATAGACCATCTGAAGACCAAGAGAGAACGATTCCTCCAGCAAACTTCAATTCAGAA 2321
Db 2403 AGACACTAGGTCAACAACTCGAGTCAACCACTCGTCCCAACACAGAACTATTTCTCA 2462
QY 2322 AGAATCTTAAGCGTCGCAAG---GGCTACATACGAACACCGGCTGTTGTTAAGGAA 2378
Db 2463 AGAAGACTTCTAAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2440
QY 2379 AACGAACCTTGAAATAAACCACCAAAA 2405
Db 2523 AGAAGACCTTAAGAGAGCTGTTTAA 2549

RESULT 5

ADD94783

ID ADD94783 standard; cDNA; 2481 BP.

XX

AC ADD94783;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human SIMP cDNA sequence.

XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;
major histocompatibility complex; human leukocyte antigen; HLA;

KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer; cancer;
KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003054008-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-CA001967.
XX
PR 20-DEC-2001; 2001US-00028384.
XX
PA (COMP-) COMPATIGENE INC.
XX
PI Perreault C, McBride K;
XX
DR WPI; 2003-559122/52.
XX
DR P-PSDB; ADD94784.
XX
XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
PT or breast cancer, or for suppressing an immune response in an autoimmune
PT disease.
XX
PS Claim 6; SEQ ID NO 1; 66bp; English.
XX
CC This invention relates to a novel isolated or purified human protein,
CC termed source of immunodominant major histocompatibility complex (MHC) -
CC associated peptide (SIMP), which is expressed ubiquitously in human
CC cells, where the protein has the potential of generating several protein
CC fragments binding with high affinity to a human leukocyte antigen (HLA)
CC molecule. The invention may allow development of therapeutics with
CC cytostatic or immunosuppressive activity or provide sequences useful for
CC antisense therapy or gene therapy. The source of immunodominant MHC-
CC associated peptide (SIMP) nucleic acids, proteins and fragments are
CC useful for diagnosing and treating cancers, for example lung cancer,
CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
CC proteins are also useful for modulating an immune response. Decreasing
CC lymphoid cell proliferation is useful for suppressing an immune response
CC responsible for an autoimmune disease or a transplant rejection. The
CC present sequence is that of the human SIMP cDNA which is related to the
CC invention. Note: This sequence does not appear in the specification but
CC was obtained by the indexer from GenBank.
XX
SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;
Query Match 39.5%; Score 955.8; DB 10; Length 2481;
Best Local Similarity 64.8%; Pred. No. 1.5e-242;
Matches 1489; Conservative 0; Mismatches 752; Indels 18; Gaps 4;
QY 102 AGTGGGCTGGGTACAGCAGCCTAATACCTTCGCCATCCCTGCTAATCGCTGGCTGGCG 161
DB 191 AGCGGCTGGGTGGGAGTGGCTTCTCTCTCCATCCATCCCTCTCCGCTGGCTGGCG 250
QY 162 GATTTCTCTCGCTTCGCGTCAATCGTTTCGAGTCGATATCCATGAGTTGATC 221
DB 251 GCTTCAGCTTCGCGCTCTTTCGCGTCAATCGCTTCGAGTCGATATCCATGAGTTGATC 310
QY 222 CGTGGTTCAACTACCGGCCACCGCTTACATGGTGCAGATGGTTGATCAACTTCCTCA 281
DB 311 CGTGGTTCAACTACCGGCCACCGCTTACATGGTGCAGATGGTTGATCAACTTCCTCA 370
QY 282 ACTGGTTTCAGCAGCGGCTGATCGCTTCGCGAGGATTCGCGGCTACCGCTATC 341
DB 371 ATTGGTTTCAGCAGCGGCTGATCGCTTCGCGAGGATTCGCGGCTACCGCTATC 430
QY 342 CCGGCTGATGATTACGCTCCGGGGGAATCAATGGCTGTCGACGCTACTCAACATACCGG 401

431 CAGGGTTGATGATAACCGCTGGCCCTATTTCATGGATTTTAAATACATTGAACATACTG 490
QY 402 TCATATTCGTCACATCTGCGTGTTCCTCGGCGCGGATCTTCAGTGGCTGACCTCCATCT 461
DB 491 TCCATAAGACGATGATGTTGTTCTTGCACCACTTTTACCGGCTTACATCTATAT 550
QY 462 CCACCTACCTGCTGACCAAGGAGTGTGCTCGGCGCGCGCTCTTCGCGCGGCTG 521
DB 551 CTACTTCTGCTTCAAGAGAACTTTGGAACCAAGGAGGAGGACTTTTACCTGCTGTT 610
QY 522 TCATCGCCATCGCTGCTGCTACATCAGTAGTGTGCTGCTGATCTGATGATGAGG 581
DB 611 TTATTCCTATTGACCGGCTACATATCTCGGTGAGTAGCTGGATCTTTGATATGAG 670
QY 582 GCATTGCCATATTCGCCCTGCACTTCACTTCTCTGTTGGTGGCTGCTGAGTGAAGCTG 641
DB 671 GCATTGCTATTTTGCACCTTCAGTTTCACTACTATTTATGGTAAATCTGTAATACTG 730
QY 642 GATCGGTTCTGCTGCGCGCGGCTTTGCTCTACTTCTACATGCTGCTGCGCTGGG 701
DB 731 GGTCACTTTTGGCAATGCTGCTCTTATCTCTATTTCTATGCTCTGCTGCTGG 790
QY 702 GTGGCTACGTTGCTATCATCACTGATPACCCCTGCACTGCTCTGCTACTGCTCATTA 761
DB 791 GTGGTTATGTTTATCATCACTTATTCACCTGCTGATGATTTGCTGTTGTTACTGATG 850
QY 762 GCAGGTACTCGCGGCTGCTGCTGACCACTACGACCTTCTACATCTCTGGAAGCTGCT 821
DB 851 AGGATGATGAGCAAGAGCTTACATAGCATATAGCATTTTCTCATTTGCTGGTTAATAT 910
QY 822 TCTCCATGAGATCCCTTCGCTGGGATTCACACCGATACGACCAAGTGAACATAGCTG 881
DB 911 TATCAATGAGATACCTTTTGGGATTCACGCAATCAGAAAGTGAACATAGCTGCGAG 970
QY 882 CGCTGGGCTGTTGCTGCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 941
DB 971 CTGAGGTGCTTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1030
QY 942 TGTGCGCAACGAGTTCGGAAGCTGTTTCATGCTGCGGAGTTCCTGCTGGGCTGCTGG 1001
DB 1031 TAACAAAACAGAGTTCAGACCTTTTCTTTTGGGTGTATCATAGCTGCTGCTGCTGCT 1090
QY 1002 TCTTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1061
DB 1091 TGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1150
QY 1062 TCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
DB 1151 TTTATTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1210
QY 1122 CGGAGCATCAGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1181
DB 1211 CTGAGCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1270
QY 1182 CTTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1241
DB 1271 CTTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1330
QY 1242 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1301
DB 1331 CTCTATATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1390
QY 1302 TACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361
DB 1391 TCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1450
QY 1362 TGAAGAGGATTCGCTGAGGAAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 1421
DB 1451 TGGGGATCA-----CATGAAGGGAATCCACCTGTGGAG-----GACAGCAGTG 1498
QY 1422 AGCTGAGGATTCATTTAGAAAGAGAGCTGTCAGCAAGGCTGGCAAGCTGAGCATC 1481

| | | |
|------|--|------|
| 1499 | ATGAGGATGACAAAGAAACCAAGGAAATTTGTATGATTAAGCGCAGGTAAAGTGAAGAAAC | 1555 |
| 1482 | GTACTAAGCATGATGCCOACGAGGATATCGCGTTCAGCTTCCAACCTGAAGAGTATTTGTTA | 1541 |
| 1559 | ATGCAACTGAACAGGAAAAAACTGAAGAGGATTTAGGCCCTAATAATAAAAAGCATTTGTCA | 1618 |
| 1542 | TTTTTGGCGTTCTTAATGCTGTTCAGTATGTTTCGCTGTCCACGTGCACGTGGGTGACCAAGCA | 1601 |
| 1619 | CCATGTTGATGCTGATGCTATTGTAATGAATGTTTCTGTCCACCTGTACCTCTGGGTCAACAAGCA | 1678 |
| 1602 | ATGCGCTACTCCAGTCCCTCCATGTTCTTTGGCTTTTCCACAACAGTCAAGATGGATCCCGCA | 1661 |
| 1679 | ATGCGCTACTCTAGTCCAAAGTGTAGTCTCTGGCTCATACAA--TTCATGATGGCACCCAGGA | 1735 |
| 1662 | ACATTTTACAGCATTTTCAGAGAGCTTACTACTGCGTTCGCGAGACACTCCCGATGATG | 1721 |
| 1736 | ATATCTTTAGATGATTTTAGAGACCTTACTCTTTTGGCTAAGGCAAAATACAGATGAACATG | 1795 |
| 1722 | CTCGCGTTATGTCCTTTGGTGGATTACGGATACCAGATAGCGGGAATGGCAACACAGAAACGA | 1781 |
| 1796 | CACGAGTAATGTCTTTGGTGGGATTTATGGCTATCAGATAGCTCGGAATGGCTAATAGAACTA | 1855 |
| 1782 | CGCTAGTCGGATAATATACGTGGACAAATAGTCACATAGCGCTGGTTGGCAGGCAATGT | 1841 |
| 1856 | CGTTCGTGGATAATAACACCTGGGAATAACAGCCACATAGCATCTGTGGGAAAAGCTATGT | 1915 |
| 1842 | CTTCAACCGAGGAGAAAGTCCCTACGAATTTATGACATCTCTTCACGTGGCACTACGTTTTGG | 1901 |
| 1916 | CTTCTAATGAACACGACGCGCTATAAATCATGAGGACTCTAGATCTAGATTATGTTTTGG | 1975 |
| 1902 | TGATCTTTGGCGGTGTGATCGGCTATTTCTGGCGATGATATCAACAAGTTTCTGTGGATGG | 1961 |
| 1976 | TTATTTTGGAGGGGTATTATGGCTATTTCTGTGATGATATCAACAAATTTCTCTGGATGG | 2035 |
| 1962 | TCGGAAATTCCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTACCGACC | 2021 |
| 2036 | TTAGATAGCTGAAGGAGAACTCCCAAGACATTCGGGAAGTGACTATTTTACCCAC | 2095 |
| 2022 | GGGTGAATTCAGGGTAGATGCCGAAGTGCTCCGGCCCTGCTCAACTGCCTTATGTACA | 2081 |
| 2096 | AGGAGAAATTCGGTGTAGACAAGACAGGATCCCTACTTTGTGTAATTCGCTTATGTATA | 2155 |
| 2082 | AATTAAGTCTACTACAGATTCGGGGAATTTGAAGTTGGCACTACAGAGGTCCTCATCTGGATG | 2141 |
| 2156 | AAATGTCATCTACAGATTTGGAGAAATGCAGCTGGATTTTCGTACACCCACGGTTTG | 2215 |
| 2142 | ATCGCACACGTAAACCGCTCATTTGGGAATTAAGGACTTCGATCTGAACCTCTGGAGAGG | 2201 |
| 2216 | ACCGAACACGTAATGCTGAGATTGGAAATTAAGGACATTAATATCAACATTTGGAAGAG | 2275 |
| 2202 | CCTACACACACAGAACACTGGCTCTTCGCACTCTATAGGGTGAAGAAGCGGATGAGTTCA | 2261 |
| 2276 | CTTTTACATCAGAACACTGGCTTTGAGGATATATAAAGTAAAGCACTGATTAACAGG | 2335 |
| 2262 | ATAGACCACTACTGAAGACCAAGGAGAGACGATTCCTCCGACAACTTCATTTTCGAGAA | 2321 |
| 2336 | AGACATTTAGATCACAAAACCTCGAGTCAACCAACTTTTCCCAAAAACAGAAAGTATTTGTCAA | 2395 |
| 2322 | AGAACTCTAAGCGTCGAAG--GGCTACATACAGAAACCGACCGGTGCTGTTAAGGAA | 2378 |
| 2396 | AGAGACTACCAAAAGGAGCGTGGCTGCATTTAAAATTAAGCTGGTTTTTAAGAAAGGCA | 2455 |
| 2379 | AACGAACCTTGAATAAAC | 2397 |
| 2456 | AGAAAATATCTAAGAGAC | 2474 |

RESIST. 6

RESULT 6
ABT.02795

ABL02795
ID ABL02795 standard: cDNA: 2855 BP.

XX
DI
LBY

AA ABL02795:

AC XX
ADL

| | | | |
|----|---|--|-----|
| DT | 26-MAR-2002 | (first entry) | |
| XX | Drosophila melanogaster | expressed polynucleotide SEQ ID NO 2867. | |
| XX | Drosophila | developmental biology; cell signalling; insecticide; | |
| XX | pharmaceutical; gene; ss. | | |
| XX | Drosophila melanogaster. | | |
| OS | | | |
| XX | WO200171042-A2. | | |
| XX | | | |
| XX | 27-SEP-2001. | | |
| XX | | | |
| XX | 23-MAR-2001; 2001WO-US009231. | | |
| XX | | | |
| XX | 23-MAR-2000; 2000US-0191637P. | | |
| PR | | | |
| XX | 11-JUL-2000; 2000US-00614150. | | |
| XX | | | |
| XX | (PEKE) PE CORP NY. | | |
| PA | | | |
| XX | | | |
| XX | Venter JC, Adams M, Li PWD, Myers EW; | | |
| PI | | | |
| XX | | | |
| XX | WPI; 2001-656860/75. | | |
| DR | | | |
| XX | P-PSDB; ABB58692. | | |
| XX | | | |
| DR | | | |
| XX | New isolated nucleic acid detection reagent for detecting 1000 or more | | |
| PT | genes from Drosophila and for elucidating cell signaling and cell-cell | | |
| PT | interactions. | | |
| XX | | | |
| XX | Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English. | | |
| PS | | | |
| XX | | | |
| XX | The invention relates to an isolated nucleic acid detection reagent | | |
| CC | capable of detecting 1000 or more genes from Drosophila. The invention is | | |
| CC | useful in developmental biology and in elucidating cell signalling and | | |
| CC | cell-cell interactions in higher eukaryotes for the development of | | |
| CC | insecticides, therapeutics and pharmaceutical drugs. The invention | | |
| CC | discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA | | |
| CC | sequences (ABU101840-ABU16175) and the encoded proteins (ABB57737- | | |
| CC | ABB72072). The sequence data for this patent did not form part of the | | |
| CC | printed specification, but was obtained in electronic format directly | | |
| CC | from WIPO at ftp.wipo.int/pub/published_pct_sequences | | |
| XX | | | |
| XX | Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other; | | |
| SQ | | | |
| | Query Match | 30.3%; Score 733.4; DB 4; Length 2855; | |
| | Best Local Similarity | 61.0%; Fred. No. 1.5e-183; | |
| | Matches 1302; Conservative | 0; Mismatches 761; Indels 72; Gaps 4; | |
| QY | 121 | CCTATACACCTTCGCCCATCTGCTTAATCGCTGGCTGGCGGATTTCTCTCGCCTCTT | 180 |
| Db | 466 | CTGTGTCAAGCTGCCATTCTCATCTCGACGGGTTTATCATTTGCCACACGCTGTT | 525 |
| QY | 181 | CGCCGTATCCGTTTCAGTTCGATTTATTCATGAGTTTGATTCGGTTCACATACCGGC | 240 |
| Db | 526 | CTCTGTGTCGGATTCCGAAAGCGTAAATCCATGAGTTCGATTCGTTCAACTACCGCAC | 595 |
| QY | 241 | CACCGCCTACATGGTGCAGAAAGGTTGGTACAACTTCTCAACTGGTTCGACGAGCGCG | 300 |
| Db | 586 | CACGCGGTTTCTGGCGGAGCAGGGCTTTTACAGTTCCACAACTGGTTCGATGACCGCG | 645 |
| QY | 301 | ATGGTATCCGCTCCGAGGATTTGGGCGGTACCGTCTATCCCGCCTGATGATTAGTC | 360 |
| Db | 646 | CTGGTATCCCTGGCGCGCATCATCGCGCGCACCATCATCCCGGGCTGATGCTCACCTC | 705 |
| QY | 361 | CGGCGGGAATCCATTGGGTGTGTGACAGTACTCAACATACCGGTCCATATTCGTGACATCTG | 420 |
| Db | 706 | GGCGGCGCTGTACCGCCTGATGTGGTGCTCAATGTGACCATCGACATACGGAACGTGTG | 765 |
| QY | 421 | CGTGTTCCTGGCGCGGATCTTCAGTGGCGCTGACCTCCATCTCCAGCTACCTGCTGACCAA | 480 |
| Db | 766 | CGTCTTCTGTGGCGGCTTCTTCTTCTCGTGTGACCGGTGGTGA CTTACGCCCTTCACAAA | 825 |
| QY | 481 | GGAGCTGTGGTTCGCGGCGCGCGGCTCTTTCGCGCGGACGCTTCATTCGCGCATCTGTCGCTGG | 540 |

826 GGAGATACACAGCACTGGAGCTGGAGCTGGTGGCGCGCTTTGATATCCATCGTTCCCGG 885
541 CTACATCAGTAGGTGGTGGCTGGATCGTAGCAATAGAGGGCATGGCCATATTCGCGCT 600
886 GTATATCTCTCGATCCGTGGCGGATCGTAGCAATAGAGCATCGCCATTTCTGCAAT 945
601 CGAGTTCACTATCTCTGGTGGTGGCTCGTAGAGACTGGATCCGTTCTCTGTCGCG 660
946 GCTCTTCACTACTATTTGTGGATCAAGCGGTAAAGACGGGACGATCTTTTGTGCGC 1005
661 CGAGCGCGCTTTGTCTTACTTCTACATGTGTCCGCTGGGTGGCTAGCTGTTCATCAT 720
1006 TATGTCGCGATTTGGCTTACTTCTATATGGTCTCTCTGGTGGTGGCTATGTTCTCTGAT 1065
721 CAACCTGATACCTCGACGCTTCTGATCTGCTGATGATGAGGAGTATCGCGCGTCT 780
1066 TAACCTAATCCGCTGACGCTGCTGGCGCTGATGATGATCAGCGGACGTTTCTCGCAGGAT 1125
781 GCTGACAGCTACAGCACTTCTACATCTCTGGAGTCTGTTCTCCATCGATCCCTT 840
1126 CTACATAGCATACAGCAGCTTACTACTGCTGGCACCATTCTGTCGATCGATCTGTT 1185
841 CGTGGATTTCAACCGATACGACAGTGAACATGCTGGTGGCTGGAGTGTGTTGCT 900
1186 TGTGGGATTTCAACCGATACGACAGTGAACATGCTGGCAGTGGGACCTTTGGGCT 1245
901 CTTTATGCGCTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 960
1246 GTGCCAGTTACAGCTTCTGTCGATCTGCTGCTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1305
961 GAAGTGTTCATCTGGCGGATGCTGCTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1020
1306 TCTGCTCTTCAAGAGCTTGGTCTTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
1021 GCTCAGCATGCTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1080
1366 GCTCAGCTTACCGGGAAGTCTCCCTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1425
1081 TGGTACGCGAAGTTCACATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1426 ATCTATGCGAAGTTCACATTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1485
1141 TTGTTCTGCTTCTTCTTCTGATCTGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
1486 CTGCTGCTTCTTCTTCTGATCTGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1545
1201 GTACTGATCAAGCAGATCAACGACGAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCT 1260
1546 TTTCTGCTTTTCCAGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1605
1261 GCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1606 TATTTATTTCCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1665
1321 GCTGCGGAGTGGCTTTTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
1666 GCTATCCGGAATGCCATTTTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1719
1381 CGGAATGGGACAGCCATAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
1720 -----TGGCGGAGCTCATCGAAGCAGGA 1744
1441 GAAGAGACGCTGTACGACAGGCTGCAAGCTGAGAGCATCTGTAAGCATGTATGCCA 1500
1745 GTAGAAAGC-----AAGCGGACGACCAAGAGCTGGAACA 1779
1501 GCAGGATCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1780 GCAGCGGAGGCTGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1839
1561 GTTATGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620

1840 GCTTATAGTGTACACGCTGCACTGCACTCGGTGCTCACTCGGAGGCTTACTCTCGCCAG 1899
1621 CATTTGCTTGGCTTTCCACACAGTCAAGATGGATCCCGCAACATTTTAGACGATTTT 1680
1900 TATTTGTTGAGTGC---CAGTCCGACGATGGGCGGCGCATATTTTCGATGACTTCG 1956
1681 AGAGGCTTACTTGGCTTTCCAGAACACTGCGCGATGCTGCTGCTGCTGCTGCTGCTGCT 1740
1957 CGAGGCTTACTTGGCTGCGATGACACTCCGAGGACGCTGCGATAATGTTCTCGT 2016
1741 GATTTACGATACACGAGTAGCGGGAATGCAAAACAGAACGACGCTAGTGGATATATATAC 1800
2017 GGAATACGCTACACGATTAACGCGCATGCGCAATCGGACGATATTTGGTGAATACATAC 2076
1801 GTGGAAATAGTGTACATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGATC 1860
2077 TTGGAAACACACATATATCGCGCTTGGCGAGCGATGCGCTTTCGGAGGAGAAAGC 2136
1861 CTACGAAATATGACATCTTTCGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
2137 CTACGAGATATAGGAGGACTGATGTGACTAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 2196
1921 CGCTTATCTGGGATGATATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1977
2197 TGGCTTACTCATCGGACGATATCAACAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2256
1978 AGAGCATCCAGGACATTAAGGAAAGGATTTACTTTACCGACCGGCTGAAATTCAGGGT 2037
2257 GATGCTGCTGCGCATCCGCGAAAGGACTTACTATCGGCAACGAGAGTTCCGAGT 2316
2038 AGATCCGAGGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2097
2317 GGAACAGGAGGCTCACCACACTCTCAATTTGTTGATGCTGCTGCTGCTGCTGCTGCTGCT 2376
2098 ATTTCGGGAAATGAAAGTTGGACTACAGAGGCTTCCATCTGATATGATCGCACACGTAACGC 2157
2377 CTTTGGGCAATGTACCGAAGGTTGCGAAAGCGGCTACCATCGAGTTCTGCTGCTGCTGCT 2436
2158 GGTATGGAATAGGACTTGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2217
2437 CGAGATCGGCAACAGGACTTTGAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2496
2218 CTGCT 2252
2497 CTGCT 2531

RESULT 7

ABL18224/c

ID ABL18224 standard; DNA; 2953 BP.

XX ABL18224;

XX AC

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Drosophila melanogaster genomic polynucleotide SEQ ID NO 6145.
Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical; gene; ds.

Drosophila melanogaster.

WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX XX

XX XX

XX XX

XX XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 6145; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention of
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABBS70721). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2953 BP; 924 A; 628 C; 648 G; 753 T; 0 U; 0 Other;

Query Match 28.8%; Score 696; DB 4; Length 2953;
Best Local Similarity 100.0%; Pred. No. 1.3e-173;
Matches 696; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAAGCGAAGATGTGTGTTGCTATTTTCAGATCGGTTATATTTTCGAGTTACTGGCTG 60
DB 696 TCTAAGCGAAGATGTGTGTTGCTATTTTCAGATCGGTTATATTTTCGAGTTACTGGCTG 637
QY 61 GAATGGGACATGATCGGACCGACGAGATCTGACAGAGGCTGCTGCTACACAG 120
DB 636 GAATGGGACATGATCGGACCGACGAGATCTGACAGAGGCTGCTGCTACACAG 577
QY 121 CCTAATACCTTCGGCATCTCTAATCGCTGCTGCGCGGATTTTCCTCTCGCCCTTT 180
DB 576 CTAATACCTTCGGCATCTCTAATCGCTGCTGCGCGGATTTTCCTCTCGCCCTTT 517
QY 181 CGCGTCATCGCTTCGAGTGAATATCCATGAGTTTGTATCGGTGTTCACTACCGGCG 240
DB 516 CGCGTCATCGCTTCGAGTGAATATCCATGAGTTTGTATCGGTGTTCACTACCGGCG 457
QY 241 CACCGCTACATGTTGCGAATGTTTGTGTACAACTTCCTCAACTGGTTTCGACAGCGCGC 300
DB 456 CACCGCTACATGTTGCGAATGTTTGTGTACAACTTCCTCAACTGGTTTCGACAGCGCGC 397
QY 301 ATGGTATCGCTCGGACGAGATGTGGGCTACCGTCTATCCGCGCTGATGATACGTC 360
DB 396 ATGGTATCGCTCGGACGAGATGTGGGCTACCGTCTATCCGCGCTGATGATACGTC 337
QY 361 CGGCGGAATCCATTGGCTGCTGACGTACTCAACATACCGGTCCATATTCGTGACATCTG 420
DB 336 CGGCGGAATCCATTGGCTGCTGACGTACTCAACATACCGGTCCATATTCGTGACATCTG 277
QY 421 CGTGTCTCGGCGGATCTTCACTGAGTGGCTGACCTGATCTCCACCTACCTGTCACAA 480
DB 276 CGTGTCTCGGCGGATCTTCACTGAGTGGCTGACCTGATCTCCACCTACCTGTCACAA 217
QY 481 GGAGCTGTGGTTCGCGGGGCGCGGCTCTTTCGCGCGGACGTTCTGCGCATCGTGGCTGG 540
DB 216 GGAGCTGTGGTTCGCGGGGCGCGGCTCTTTCGCGCGGACGTTCTGCGCATCGTGGCTGG 157
QY 541 CTACATCAGTAGTGGTGGCTGGATCTGATGATGATGATGATGATGATGATGATGATGATG 600
DB 156 CTACATCAGTAGTGGTGGCTGGATCTGATGATGATGATGATGATGATGATGATGATGATG 97
QY 601 GCAGTTTCACTACTTCTCTGCGGTGGCTTCAGTGAAGACTGGATCCGCTGTTCTGGTGGCG 660
DB 96 GCAGTTTCACTACTTCTCTGCGGTGGCTTCAGTGAAGACTGGATCCGCTGTTCTGGTGGCG 37
QY 661 CGGACGCGCTTTGCTCTCTTCTACATGATGTTCCGC 696

Db 36 CGCAGCGCTTGTCTACTTCTTCTACATGTTGTCGC 1

RESULT 8

ABT20789
ID ABT20789 standard; DNA; 2232 BP.
XX
XX AC ABT20789;
XX
XX DT 16-APR-2003 (first entry)
XX
XX DE Aspergillus fumigatus essential gene #3147.
XX
XX KW Fungicide; cytosolic; essential gene; Aspergillus fumigatus; infection;
XX cancer; contamination; biofilm; antibody; immune response; ds.
XX
XX OS Aspergillus fumigatus.
XX
XX PN WO200286090-A2.
XX
XX PD 31-OCT-2002.
XX
XX PF 23-APR-2002; 2002WO-US013142.
XX
XX PR 23-APR-2001; 2001US-0285697P.
XX
XX PR 27-APR-2001; 2001US-0287066P.
XX
XX PR 05-JUN-2001; 2001US-0295890P.
XX
XX PR 09-JUL-2001; 2001US-0303899P.
XX
XX PR 31-AUG-2001; 2001US-0316362P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
PT
PS Disclosure; Page; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterization, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention
XX
XX SQ Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;

Query Match 26.3%; Score 635.4; DB 8; Length 2232;
Best Local Similarity 58.9%; Pred. No. 1.4e-157;

| Matches 1248; Conservative 0; Mismatches 796; Indels 75; Gaps 6; | |
|--|--|
| Qy | 137 ATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGGCTCTTGGCGTATCGGTTTC 196 |
| Dd | |
| Qy | 76 ATCTGCTAATCGCTGGCTGGCGGATTTTCCTCTCGGCTCTTGGCGTATCGGTTTC 135 |
| Dd | |
| Qy | 197 GAGTCGATTTATCCATGATTTGATCCGCTGTTCAACTACCGGGCCACCGCTTACATGGTG 256 |
| Dd | |
| Qy | 136 GAGAGTATCATCCAGGAAGTGGCCGCTGTTCAACTTCGAGCAACAAATACCTAGTA 195 |
| Dd | |
| Qy | 257 CAGATGTTGGTACATCTTCCATCGCTGTTGAGCGGCGCATGTTATCGCTCGGC 316 |
| Dd | |
| Qy | 196 CAGATGTTGTTATGATCTTTGGGATTTGGATTTGGATGACGGAACATGGCATCTCTGGGA 255 |
| Dd | |
| Qy | 317 AGGATTTGGCGGTACCGTCTATCCCGGCTGATGATTAACGTCGCGGGAATCCATTGG 376 |
| Dd | |
| Qy | 256 CGTGTACCGGTGGCAGTTATATCCCGGTCTCATGGTACGAGCGGCTGATCAGCAT 315 |
| Dd | |
| Qy | 377 CTGCTGCACGTACTCAACATACCGTCCATATTCGTGACATCTGGTGTTCGTGGCGG 436 |
| Dd | |
| Qy | 316 ATCTTGGGATTTCTATCTATCCCGCTGCAATTCGCAACATCTGCTCTCTGGGGA 375 |
| Dd | |
| Qy | 437 ATCTTCACTGCGCTGACCTCCATCTCCACCTACTGCTGACCAAGAGCTGTGGTCCGG 496 |
| Dd | |
| Qy | 376 GGAATCTCCGGCTGACGTGATTTGGCAATGTAATCTGACATCCGAGATGTCTCTCG 435 |
| Dd | |
| Qy | 497 GGC---GCCGCTCTTGGCGGCGAGCTTCATCGCATCGCTGCTGCTACATCAGTAGG 553 |
| Dd | |
| Qy | 436 CCATCTGAGGCTTCTTGGAGAGCTTTTATGGGAATCGCCCTGGTACATCTCCGA 495 |
| Dd | |
| Qy | 554 TCGETGGCTGATCGTACGTAACGAGGCAATTCGCATATTCGCCCTGACGATTCACCTAC 613 |
| Dd | |
| Qy | 496 TCAGTGTGGAAGTACGTAACGAGAGCAATTTGCCATCTTTCTGCTGTGTTCACATTC 555 |
| Dd | |
| Qy | 614 TTCCTGTGGTGGCTCAGTGAAGCTGATCGTGTCTGCTGGCGGCGAGCGCTTTG 673 |
| Dd | |
| Qy | 556 TTCTATGGAACAAGGTGTCAAAATGGGTCTATCATGTGGGAGCGCTGACCGCACTA 615 |
| Dd | |
| Qy | 674 TCCTACTTTACATCTGCTGCGCTGGGTGGCTAGTGTTCATCATCACTGATACCC 733 |
| Dd | |
| Qy | 616 TTCTACGGCTACATGTTGTCGGCATGGGTGGGTATGTTCTTACATACGAGCTGATCC 675 |
| Dd | |
| Qy | 734 CTGACGCTTCTGCTACTGCTCATATGGGAGGCTACTCGCGGCTGCTGACGAGCTAC 793 |
| Dd | |
| Qy | 676 CTGACGCTTTTGTCTCTGTCGATGGGTAGATACGACCTGCACTACATTAAGCTAT 735 |
| Dd | |
| Qy | 794 AGCACCTTTACATCTCCGGGAGTCTGTTCTCCATGACATCCCTTCGTGGGATTCGA 853 |
| Dd | |
| Qy | 736 ACCATGTTATGCGTGGGAGCTTTGGCTAGCATGCAGATTCCTTCGTGGATTTTG 795 |
| Dd | |
| Qy | 854 CCGATAGCACGAGTGAACATGGCTGCTGGGAGTGTGTGCTCTTATGCGGCTG 913 |
| Dd | |
| Qy | 796 CCTATCGAACAACGACCACTATGTCGCTTGGGTGCTTTCGCGCTGCTTCAGCTTGTG 855 |
| Dd | |
| Qy | 914 GCCACCTTGGCCATTTGCACTGCTGCTGCGCAACGAGTTCGGAAGCTGTTCATC 973 |
| Dd | |
| Qy | 856 GCCTTGGCCGAGTTTGTCCGAGCTTCGTTCCAAAGCAGTTCAGAGACTTCTGACC 915 |
| Dd | |
| Qy | 974 GTGCGGAGTGTGTTGGGCTGGGCTTGTGTGGCGCTGCTGCTGCTACCATGCTG 1033 |
| Dd | |
| Qy | 916 GCCATGATCTTACATCTTGGCTGCTGGTTCGCTTGGGCTAGTGTCTGACTGTGAGC 975 |
| Dd | |
| Qy | 1034 GCGTGTGGCCCGCTGGAGTGAAGCTTCTACTGCTGTGGGATACGCTACGCCAAG 1093 |
| Dd | |
| Qy | 976 GGAGTGTACGCTCTTGGAGCGGCGGATTTCTACTCTTGTGGGACACTGGCTATGCCAAA 1035 |
| Dd | |
| Qy | 1094 ATCCACATTCCTCATTTGATCGTGGAGCATCAGCCACCACTTGTTCGTCTC 1153 |
| Dd | |
| Qy | 1036 ATCCACATTCCTCATTTGATCTGAGTCTCGAACAACGCGCCACCGCTTGGCCAGCTTC 1095 |
| Dd | |
| Qy | 1154 TTCTTTGATCTGCAATCTTGGTGTGCGCTTCCGAGTGGAGTGTGGTACTGATCAAG 1213 |
| Dd | |
| Qy | 1096 TTCTTCGATCTGAATCTTCTGATCTGGCTTTTCCCGGAGGTGTCTACATGTGCTTCGT 1155 |
| Dd | |

| | |
|----|--|
| Qy | 1214 CAGATCAACGACGAGCGGCTTTTCGTGGTGTGTAGGCATCACTGCGGTTTACTTCCGT 1273 |
| Dd | |
| Qy | 1156 GACCTCAAGAGAGAGCATGTCTTCGTCAATATCTACTCGGCTCTTGGAGCTACTTCCGC 1215 |
| Dd | |
| Qy | 1274 GGTGCTGATCGTTCGTGTGATGTGACCTCAACGCGGTGGTGTGATGCTGCGCGAGTG 1333 |
| Dd | |
| Qy | 1216 GGTGTTATGTTCCGACTAATGCTGACCTTGACCCCTATTGTGTGTGTGGCGCTGCTCTG 1275 |
| Dd | |
| Qy | 1334 GCCTTTTCGGGACTGTGTGATGTGTTCTGCAAGAGGATTCGTCTAAGCGAATGGGCACA 1393 |
| Dd | |
| Qy | 1276 GCCTGTCTTCATTTCTGACACCTATATG----- 1305 |
| Dd | |
| Qy | 1394 GCCTAAGCGGACGACCGAAGTGGATGAAGCTGAGGATTCATTTGAGAGAGAGACGCTG 1453 |
| Dd | |
| Qy | 1306 GCGACTACCTCCCGACACCGAGCTCTGAAG-----G 1338 |
| Dd | |
| Qy | 1454 TAGCAAGGCTGGCAAGCTGAAGCATGCTACTAAGCATGATGCCAGGAGTACTGGC 1513 |
| Dd | |
| Qy | 1339 AAGAGATGAAGACTGCTCTTCACACCTCTTCGTGCTCGAAGCCCAATTTGGA 1398 |
| Dd | |
| Qy | 1514 GTCAGCTCCAACTG---AAGAGTATGTTATTTTGGCGCTTAAATGCTGTGATGATG 1570 |
| Dd | |
| Qy | 1399 ATCAGCTCCATGTTTCTAAGATTATAGTACGCGCTCTGTTGCTGCTACTCTCTCTG 1458 |
| Dd | |
| Qy | 1571 TTCGCTGCTCCACTGACGCTGGGTGACCAATGCTACTCCAGTCCCTCCATTTGCTTG 1630 |
| Dd | |
| Qy | 1459 TTTGTTGGCATGTGACCTGGGTTACATCGAATGCTACTCTTCTCTCTCGTTGCTTG 1518 |
| Dd | |
| Qy | 1631 GCTTTCCCAACAGCTCAAGATGATGCTCCGCAACATTTTACAGATTTTACAGAGGCTTAC 1690 |
| Dd | |
| Qy | 1519 GCTAGTCCGATGCT---GACGGAAGCCCAATACATCATTCAGGATTTCTGAGGCTTAC 1575 |
| Dd | |
| Qy | 1691 TACTGCTTTCCGAGAACACTGCGCATGATGCTGCGTTATGTTCTGTTGGGATTTACGGA 1750 |
| Dd | |
| Qy | 1576 TACTGCTTCTCGAGATATCTCTCCAGAACGCAAAATCATGTCTATGTTGGGATTTG 1635 |
| Dd | |
| Qy | 1751 TACAGATAGCGGGAATGGCAAAAGAACAGACGCTGATGGAATATAATACGTGGAACAAT 1810 |
| Dd | |
| Qy | 1636 TATCAAAATCGGTGGATGCGGACCGCCCAACCTTTGTTGAACAACAACACCTGGAAACA 1695 |
| Dd | |
| Qy | 1811 AGTCACATAGCTGTTGGGCAAGCAATGCTTCAACCGAGGAGAGTCTCTACGAAT 1870 |
| Dd | |
| Qy | 1696 ACCATATTTGCTACGTTGTTGAGCGATGAGCTCACGCGAGGAGTACGCTACCCCATC 1755 |
| Dd | |
| Qy | 1871 ATGACATCTTTGAGCTGACTACGTTTGTGATGTTTGGCGGTGTGATCGGCTATTCT 1930 |
| Dd | |
| Qy | 1756 CTCCGCGAGCATGATGCTGATTTACGTTGCTGGTGTGTTTCCGCTGCTGCTAGGTTATCT 1815 |
| Dd | |
| Qy | 1931 GCGGATGATCAACAAGTTCCTGTGGATGTTCCGAATTTGCTGAGGAGAGAGCATCCCAAG 1990 |
| Dd | |
| Qy | 1816 GCGGATGATCAATTAACAATTTCTTATGATGTTCCGATCGCGAGGATTAATCGCCGAT 1875 |
| Dd | |
| Qy | 1991 GACATTAAGAAAGCGATTTACTTTACGACCGGCTGAAATTCAGGGTAGATGCGGAGGT 2050 |
| Dd | |
| Qy | 1876 GAGGTTAAAGAGCGGAGCTTTCTTACTGACGCGGTGAATATCGTGTGACGATGGAGCG 1935 |
| Dd | |
| Qy | 2051 GCTCGCGGCTCTCACTCACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGAATTG 2110 |
| Dd | |
| Qy | 1936 ACCCAACTATGGGCAACAGCTTGTGTTAATAATGTTTATTACA-----ATTTC 1986 |
| Dd | |
| Qy | 2111 AAGTTGGATCAGAGGTCCTATCTGATATGATGCAACACATTAACGCGCTCAITGGGAAT 2170 |
| Dd | |
| Qy | 1987 AACTCTCTCTTCCCGTGGGCGCAAGCTGTCGACCGGCTCGCGTGGGTCAAAATTTCCACA 2046 |
| Dd | |
| Qy | 2171 AAGGACTTCGATCTGACCTTACCTGAGGAGGCTTACACACAGAACACTGCTGCTTGTTCG 2230 |
| Dd | |
| Qy | 2047 GAAGGCTCTCAGTCTCTACATCTCGAAGAGCTTTCAGCGGAGAGTCTGATCATTCGT 2106 |
| Dd | |
| Qy | 2231 ATCTAAGGTTGAAGAGC 2249 |
| Dd | |
| Qy | 2107 ATCTAAGGTTCAAGGATC 2125 |
| Dd | |

RESULT 9

ADD94793
ID ADD94793 standard; DNA; 2472 BP.

AC ADD94793;

DT 29-JAN-2004 (first entry)

XX Human ITM1 gene sequence.

DE Human ITM1 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;

KW major histocompatibility complex; human leukocyte antigen; HLA;

KW cystostatic; immunosuppressive; antisense therapy; gene therapy; cancer;

KW lung cancer; intestine cancer; sarcoma; prostate cancer;

KW testicular cancer; breast cancer; melanomas; pancreatic cancer;

KW haematological cancer; immune response; lymphoid cell proliferation;

KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;

XX gene; ds; ITM1.

OS Homo sapiens.

XX WO2003054008-A2.

FN 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

PF 20-DEC-2001; 2001US-00028384.

PR (COMP-) COMPATIGENE INC.

PA Petreault C, McBride K;

XX WPI; 2003-559122/52.

XX P-PSDB; ADD94794.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic

acids and proteins, useful for diagnosing and treating cancer, e.g. lung

disease.

PT Disclosure; SEQ ID NO 11; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,

CC termed source of immunodominant major histocompatibility complex (MHC) -

CC associated peptide (SIMP), which is expressed ubiquitously in human

CC cells, where the protein has the potential of generating several protein

CC fragments binding with high affinity to a human leukocyte antigen (HLA)

CC molecule. The invention may allow development of therapeutics with

CC cytostatic or immunosuppressive activity or provide sequences useful for

CC antisense therapy or gene therapy. The source of immunodominant MHC-

CC associated peptide (SIMP) nucleic acids, proteins and fragments are

CC useful for diagnosing and treating cancer, for example lung cancer,

CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast

CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP

CC proteins are also useful for modulating an immune response. Decreasing

CC lymphoid cell proliferation is useful for suppressing an immune response

CC responsible for an autoimmune disease or a transplant rejection. The

CC present sequence is that of the human ITM1 gene which is related to the

CC invention. Note: This sequence does not appear in the specification but

CC was obtained by the indexer from GenBank.

XX SQ Sequence 2472 BP; 566 A; 568 G; 583 G; 755 T; 0 U; 0 Other;

Query Match

Best Local Similarity 26.2%; Score 632.8; DB 10; Length 2472;

Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;

QY 118 CACCTTAATACCTTCGCCATCTGCTAATCGCCTGGCTGGCGGATTTCTCTCGCT 177

Db 151 CACACTTTTGAAGCTTCTCATCTGTCAATGGCTGTGTATATCTCTCCACTCGCT 210

QY 178 CTTGCGCGTCAATCGGTTTGGAGTCGATTAATCCANGAGTTTGATCCCGTGGTTCAACTACCG 237

Db 211 GTTTGTGTCCTCGAGATTTGAAAGTGTGTATCCATGAGTTTGATCCCGTACTTTAAATATCG 270

QY 238 GGCACCGCCCTACATGGTGCAGAAATGGTTGGTACAACTTCCCTCAACTGGTTGGACGAGG 297

Db 271 GACTACCAAGGTTCTCGCTGAGGAGGGGTTTATATAATTCATAACTGGTTTGATGACCG 330

QY 298 CGCATGGTATCCGCTCGGAGAGATTTGGGGGATACCGTCTATCCGGGCTGATGATTAC 357

Db 331 AGCGTGGTACCCCTTTGGGAGCAATCATTTGGAGGAAACAATTTACCCAGGTTTAATGATCAC 390

QY 358 GTCCGCGGAATCCATTGGCTGTCGACGTACTCAACATACCGTCCATATTTGTCGACAT 417

Db 391 CTCTGCTGCAATCACTACCATGTAATCAATTTTCCACATCAACATGACATTCGGAATG 450

QY 418 CTGCGTGTCTCTGGCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC 477

Db 451 CTGTGTGTTCTCTGGCGCCCTCTCTTCTCTCTCTCACTCCATCGTCACGTACCTCTTAC 510

QY 478 CAAGGAGCTGTGTCGCGGGCGCGGCTCTTCCGCGCCAGCTTCAATCGCATCGTGCC 537

Db 511 CAAAGAGCTCAAGGATCGAGGGGCTGGGCTTCTTCTGCTGCCATGATTTGCTGTAGTTCC 570

QY 538 TGGCTACATCAGTAGGTGGTGGCTGGATCGTACGATAACGAGGGCAATTCATATTCGC 597

Db 571 TGGATATATCTCCGATCTGTGGCTGGCTCTTATGATTAATGAAGGATTTGCCATCTTTG 630

QY 598 CTGCGAGTACAGTACTTCTCTGGTGGCTGAGTCAAGAGAGTGAATCGGTGTTCTGGTC 657

Db 631 CATGCTACTCAGTACTCATGTGGATCAAGCAGTAAGACTGGTGGTCCATCTGTGGGC 690

QY 658 GCGCGCAGCGGCTTTGTCTCTACTTCTACATGGTGTCCGCTGGGTGCTAGCTGTTCTAT 717

Db 691 AGCTAAGTGTGCCCTTGTCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTTCTCT 750

QY 718 CATCAACCTGATACCCCTGCGAGTCTTCTGCTACTGCTCATTAATGAGCAGGTAATCGCCGC 777

Db 751 GATCAACTTAATCTCTCTCCAGCTGCTGCTGATGCTCACAGGCCGTTCTCTCACCG 810

QY 778 TCTGCTGACCACTACAGCACCTTCTACATCTCGGAGTGTGTTCTCCATGCGAGATCC 837

Db 811 GATCTATGTGGCTTACTGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 870

QY 838 CTTCTGCGGATTCGAACCGATACGACACAGTGAACACATGCTGCTGCTGCTGCTGCTGCTG 897

Db 871 CTTTGTGGGTTTCCAGCCGTCTTCTTCTCATCAGACACATGCGAGGGTTTGGGCTTTGG 930

QY 898 GCTCCTTATGGCGTGGCGCACCTTGGCCATTTGCGAGTCCGCTGCTGCTGCTGCTGCTGCTG 957

Db 931 TCTCTGCGAGATCCATGCTCTTTGGAATACCTGCGCAGCAAGTTGAATCCACAACAAT 990

QY 958 CCGGAAGCTGTTCTATCGTGGCGGATTTGCTGGTGGGCTTGGCGCTCTTTGCTGGCGCTG 1017

Db 991 TGAAGTCTTTTTCGGAGCGTCATCTCTCTGTTAGGCTTTGTCTCTTCTCACCGTGGGAGC 1050

QY 1018 GGTGCTCAGCATGCTGGCGGTTGTTGGCCCGTGGAGTGGAGCTTCTACTGCTGTGGGA 1077

Db 1051 TCTCCTCATGCTGACGAGAAAATATCTCCCTGGACGGGGGCTTCTACTCATCTGCTGGA 1110

QY 1078 TACTGGCTTACGCCAAGATCCCAATTTCCATTCATTCGCTGCTGCTGCTGCTGCTGCTGCTG 1137

Db 1111 TCCCTCTTATGCTAAGAACACATCCCAATTCATTCCTTCTGTTCTGAGCATCAGCCAC 1170

QY 1138 CACTTGTGTTCTGTTCTTCTTTGATCTGCAATCTCTGTTGCTGCTGCTGCTGCTGCTGCTG 1197

Db 1171 AACCTGCTCTCATATCTTTCACCTGCGAGTCTCTGCTCTTCTGCTCTTCTGCTGCTGCTG 1230

QY 1198 GTGCTACTGATCAAGCAGATCAACGACGAGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1257

Db 1231 CTATTACTGCTTTAGCAACCTGCTGATGCCGGAATTTTATCATCAATGATGCTGCTGCTG 1290

QY 1258 TGGGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317

1291 CAGCATGACATTTTTCAGCTGTAATGTCGCTGTAATGTCAGTGTGGCAGCTGTAATGAG 1350
 1318 CATGCTGCGGAGTGGCTTTTGGGACTGTGTGATGTTCTCTGCAAGAGGATTCGTC 1377
 1351 CATCTCTCTGTCATGAGTCTCCAGTCTG----- 1384
 1378 TAAGCGAATGGGCACAGCCATAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCAT 1437
 1385 -----TCCAC 1389
 1438 TGAGAGAGAGCGCTGTACGACAGGCTGGCAGCTGAAGCATCGTACTAAGCATGATGC 1497
 1390 ATACATGAAGATCTGGACATAGTCGCCACAGCAAGAGAGCAAGAGCAAGAGATTC 1449
 1498 CCAGCAGGATATCTGGCGTCAGCTCCAACTGAAGAGTATTGTTATTTTGGCCGTTCTAAT 1557
 1450 CACCTACCTATTAGATT-----GAAGTGGCAAGTGGATGATCTGCTATGGCTTT 1503
 1558 GCTGTGATGATGTTGCTGTCCACTGACGTGGGTGACCGACATGCTTACTCCAGTCC 1617
 1504 CTCTCTCATCACTACACTTTTCACTCAACCTGGGTGACAGTGGGCTTACTCTCTCC 1563
 1618 CTCCTATGCTTGGCTTTCCACACAGTCAAGATGATCCCGCAACATTTTAGACGATT 1677
 1564 GTCCATTGTACTATCTGCC-----GTGGTGGATGCGAGTATCATATTGATGACTT 1620
 1678 CAGAGAGGTTACTACTGCTGTTCCGAGAACACTGCGGATGATGCTCGGTTATGCTTG 1737
 1621 CCAGAGCATATATTGCTGCTGTCATTAATCTCCAGAGGATGCAAGGTCATGCTCG 1680
 1738 GTGGGATTCAGATACCATGAGCGGATGCAACAGACGACGCTAGTGGATAATA 1797
 1681 GTGGGATTTGGCTATCAGATTACAGTATGCAATGCAACCGAACAATTTAGTGGCAATA 1740
 1798 TAGTGGAAATAGTACATAGCGCTGTTGGCAAGCAATGTTTCAACCGAGGAGAA 1857
 1741 CACATGGAATAATACCATATTCTCGAGTAGGCGAGCAATGGGCTCCACAGAGGAAAA 1800
 1858 GTCTAGCAAAATATGACATCTCTTACGCTGACGCTAGCTTTTGGTGTATCTTTGGGGGT 1917
 1801 AGCCTATGATCATGAGGAGCTCGATGTGAGTATGCTGGTCAATTTTGGAGGCT 1860
 1918 GATCGGCTATCTCGGCGATGATCAACAAGTTCTGTGGATGTCGCGAATTCGTGAGG 1977
 1861 CACTGGGATTTCTCTGATGATATCAACAAGTTTCTTTGGATGTCGCGAATGGAGGAG 1920
 1978 ---AGAGCATCCCAAGACATTAAGGAAGCGTATCTTTACCGACCGGCTGAATTCAG 2034
 1921 CACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCAATCTGGGAGTCCG 1980
 2035 GGTAGATGCGGAAGGTCCTCGGCTGCTCAACTGCTTATGTAACAAATTAAGCTACTA 2094
 1981 TGTGGACCGTGAAGTTCTCCAGTGTCTCACTGCTCACTGCTCAAGATGTTACTA 2040
 2095 CAGATTCCGGGAATGAAGTTGGAATGACATGACAGGTCCATCTGATATGTCGACACGTA 2154
 2041 TCGCTTTGGACAGGTTTACACAGAACCCAGCGCTCTCCAGGCTTTGACCGGTGTCGAAA 2100
 2155 CGCGCTCATTTGGGAATTAAGGACTTCGATCTGACCTACCTGGAGAGGCTTACACACAGA 2214
 2101 TGTGAGATTGGGATTAAGACTTTGAGCTTGATGCTTGGAGAGGCTTATACACAGA 2160
 2215 ACATGCTGTTTGGCATCTATAGGTTGAAGAGCGCATGA 2256
 2161 ACATTGCTGTGAGGATATACAGGTAAAGGACCTGGATA 2202

RESULT 10
 ADH28844
 ID ADH28844 standard; DNA; 2472 BP.
 XX
 AC ADH28844;

XX 11-MAR-2004 (first entry)
 DT Human chronic myelogenous leukaemia (CML) gene marker #112.
 XX
 DE da; chronic phase chronic myelogenous leukaemia; CP-CML;
 XX blast crisis CML; BC-CML; human; chronic myelogenous leukaemia;
 KW gene marker.
 KW
 XX Homo sapiens.
 OS
 XX US2003104426-A1.
 PN
 XX 05-JUN-2003.
 PD
 XX 14-JUN-2002; 2002US-00171581.
 PF
 XX 18-JUN-2001; 2001US-0298914P.
 PR
 XX (LINS/) LINSLEY P S.
 PA (MAOM/) MAO M.
 PA (DAIH/) DAI H.
 PA (HEYI/) HE Y.
 PA (RADI/) RADICH J P.
 XX
 XX Linsley PS, Mao-M, Dai H, He Y, Radich JP;
 PI WPI; 2003-787046/74.
 XX
 DR Classifying cell sample as chronic phase chronic myelogenous leukemia or
 XX blast crisis chronic myelogenous leukemia by detecting difference in
 PT expression of genes corresponding to the markers such as X15415, U89436.
 PT
 XX Disclosure; SEQ ID NO 112; 31pp; English.
 PS
 XX The invention relates to a method of classifying a cell sample as chronic
 CC phase:chronic myelogenous leukaemia (CP-CML) or blast crisis CML (BC-
 CC CML). The method is useful for classifying a sample as CP-CML or BC-CML.
 CC The present sequence represents a human chronic myelogenous leukaemia
 CC (CML) gene marker used to distinguish blast crisis CML from chronic phase
 CC CML.
 XX
 XX Sequence 2472 BP; 566 A; 568 C; 583 G; 755 T; 0 U; 0 Other;

Query Match 26.2%; Score 632.8; DB 10; Length 2472;
 Best Local Similarity 58.2%; Pred. No. 6.9e-157;
 Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;
 QY 118 CAGCCTAATCACTCGCCATCTGCTAATCGCTGCGCTGCGCGGATTTTCTCTCGCT 177
 Db 151 CACATTTTGAAGTCTCTCATCTGTCATGCTGCTGTTATTCCTTCTCCACTCGTCT 210
 QY 178 CTTCCGCGTTCATCGTTTCGAGTCGATATCCATGAGTTTGTATCCGCTGTTCACTACCG 237
 Db 211 GTTTGCTGTCTCGAGATTTGAAAGTGTATCCATGAGTTTGTATCCGCTACTTTAATATCG 270
 QY 238 GGCCACCGCTACATGTCGAGATGTTGTCGTAACATCTTCTCAACTGGTTCGACGAGCG 297
 Db 271 GACTACCAAGTTCTGCTGCTGAGGAGGTTTATTAATTCCTAACTGTTGATGACCG 330
 QY 298 CGCATGTTATCCGCTCGGCGAGGATTTGGCGGTATCCGCTATCCCGGCTGATGATTAC 357
 Db 331 AGCTGTGTACCTTTGGGACGAATCATTTGGAGGAACAATTTTACCCAGGTTTAAATGATCAC 390
 QY 358 GTCCGGCGGATCCATTTGGCTGTCGACGACTCAACATACCGTCCATATTCGTCACAT 417
 Db 391 CTCTGCTGCAATCTACCATGTTCTCCATTTTCCATCATCACCATCGACATTCGGAATGT 450
 QY 418 CTGCTGTTCTGCGCGCGATCTTTCAGTGGCTGACCTCCATCTCCACCTTCTGCTGAC 477
 Db 451 CTGTGTGTTCTGCGCCCTCTCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 510
 QY 478 CAAGAGGCTGTGTCTCGGCGGCGGCGGCTTCTTCCGCCGCGGCTTCTATCCCATCGTGC 537

Db 511 CAAGAGCTCAAGGATCAGGGGCTGGCTTCCTGCTGCCATGATTCCTAGTTC 570
Qy 538 TGGCTACATCAGTAGTTCGGTGGTGGATCGTAGATAAGCAGGCAATGCCATATTCGC 597
Db 571 TGGATATATCTCCGATCTGGCTGGCTGCTATGATAATGAGGATTCCTATTTG 630
Qy 598 CTGCAAGTTCACCTACTCTCTGCTGGTGGCTCAGTGAAGACTGGATCCGTCTGCTC 657
Db 631 CATGCTACTCACCCTACTACATGAGATCAAGCAGTAAAGACTGGTTCATCTGTTGGC 690
Qy 658 GCGCCAGCAGCGCTTGTCTCTACTCTACATGGTTCGCGCTGGGGTGGCTACGTTCAT 717
Db 691 AGCTAAGTGTGCGCTTGTCTATTTCTACATGGTCTGCTCATGGGGAGTTATGTTCT 750
Qy 718 CATCAACCTGATACCCCTGCAAGCTTCCTGCTACTCTCATTTATGGCAGTACTCGCGG 777
Db 751 GATCAACTTAATTCCTCTCCACGCTCTGCTGATGCTCACAGCGCGTTCCTCACCG 810
Qy 778 TGTGCTGACCACTACAGCACTTCTACATCTCTGGGACTGCTGTCTCCATGAGATCCC 837
Db 811 GATCTATGTGGCTACTGCTACTGTTTACTGCTGGTACTATATCTTTCTAGGCAGATCTC 870
Qy 838 CTTCGTGGGATTCGAACGATACGACACAGTGAACATGCTGCTGCGTGGAGTGTGT 897
Db 871 CTTCGTGGGTTTCCAGCCTGCTCTTTCATCAGACACATGGCAGGGTTTGGGGTCTTGG 930
Qy 898 GCTCTCTTATGGCGCTGGCCACTTTCGCGCATTTGCACTGCTGCTGCGGCAACAGATT 957
Db 931 TCTCTGCCAGATCCATGCTTGTGGATTACTCTGCGCAGCAAGTTGAATCCACAACAT 990
Qy 958 CCGAAGCTGTATCTGCTGGGATTCGCTGGGATTCGCTGGGCTGCTGCTGGGCTGCT 1017
Db 991 TGAAGTCTTTTCCGAGCGTCTCTCTGCTGGAGCTTGTCTCTCTACCGTGGGAGC 1050
Qy 1018 GGTGCTCACCATGCTGGGCTGTGGCCCGCTGGAGTGGACGCTTCTACTCGCTGGGA 1077
Db 1051 TCTCTCTGCTGACAGGAAATATCTCTGCGACGGGCTTCTACTCTGCTGGA 1110
Qy 1078 TACTGGCTACGCAAGATTCACATCCCATCTGATTCGATTCGCTGCGAGCATCAGCCAC 1137
Db 1111 TCCCTCTTATGTAAGAACACATCCCATCATTCGCTGCTGCTGCTGCTGCTGCTGCT 1170
Qy 1138 CACTTGGTCTCTGCTTCTTGTGATCTGCATCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
Db 1171 AACCTGGTCTCATACTATTTGACCTGCAGCTCTCTGCTCTCATGTTTCCAGTTGCGCT 1230
Qy 1198 GTGGTACTGCATCAGCAGATCAAGCAGAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCT 1257
Db 1231 CTATTACTGCTTTAGCAACCTGCTGATGCGCGGATTTTATCATCATGATGATGCTGCTG 1290
Qy 1258 TCGGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
Db 1291 CAGCATGATCTTTTTCAGCTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350
Qy 1318 CAGCTGCGCGAGTGGCTTTTTCGGGACTGTTGGATGCTGCTGCTGCTGCTGCTGCTGCT 1377
Db 1351 CATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
Qy 1378 TAAGGAATGGGCACAGCCATAAGCGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1437
Db 1385 -----TCCAC 1389
Qy 1438 TGAGGAAGAGCCTGTAGCAGAGCTGGCAGCTGAGCATCGTACTAAGCATGATGC 1497
Db 1390 ATACATGAAGATCTGGACATAGTTCGCCAGACAAGAGAGCAGCAGCAGCAGCAGCAG 1449
Qy 1498 CCAGCAGGATCTGGCGCTCAGCTCCCACTGCAAGATGATTTGTTTGGCGCTTCTAAT 1557
Db 1450 CACCTACCTATTAAGATT-----GAAGTGGCAAGTGGGATGATCTGCTGCTGCTGCT 1503
Qy 1558 GCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617

Db 1504 CTTTCTCATCACCTACACCTTTTCATTTCAACCTGGTGACCAAGTGGGCTACTCTTCTCC 1563
Qy 1618 CTCCATTGTCTTGGCTTTCCACAAACAGTCAAGATGGATCCCGCAACATTTTAAACGATTT 1677
Db 1564 GTCCATTGTACTATCTCTCCCTC---GTGGTGGGATGGCAGTAGGATCATATTTGATGACTT 1620
Qy 1678 CAGAGAGCTTACTACTGCTTTTCGAGAACACTGCGGATGATGCTCGGCTTATGCTTGTG 1737
Db 1621 CCGAAGACATATTTTGGCTTCTGTCATATATCTCCAGAGATGCGAAGTCTATGCTCTG 1680
Qy 1738 GTGGATTACCGATACAGATAGCGGATGGCAACAGAAACAGCTAGTGGATTAATAA 1797
Db 1681 GTGGGATTTATGGCTATCAGATTTACAGCTATGGCAACCGAACAAATTTTATGTCACAATAA 1740
Qy 1798 TACGTGAAACAATAGTCACATAGCGCTGGTGGCAAGGCAATCTCTTCAACCGAGGAA 1857
Db 1741 CACATGGAATATACCCATATTTCTCGATGAGGAGGCAATGGCTCCACAGAGGAAA 1800
Qy 1858 GTCTACGAAATATGACATCTCTTGACGTGGACTAGTCTTTGGTGATCTTTGGCGGTGT 1917
Db 1801 AGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTGCTCATTTTGGAGGCT 1860
Qy 1918 GATCGCTATTTCTGGGATGATATCAACAAGTTTCTGTTGATGCTCCGAATTCCTCAGGG 1977
Db 1861 CACTGGGTATTTCTCTGATGATATCAACAAGTTTCTTGGATGCTCGGATTTGGAGGAG 1920
Qy 1978 ---AGAGCATCCCCAAGACATTAAGAAACGATTACTTTTCCGACCGCGGTGAATTCAG 2034
Db 1921 CACAGATACAGGCAACATATCAAGGAGATGACTATATCTTACTTCCAACTGGGAGTTCCG 1980
Qy 2035 GGTAGATGCGGAGGCTCTCGGCTTCTCAACTGCTTATGATCAAAATTAAGTACTA 2094
Db 1981 TGTGACCGTGAAGTTCTCCAGTCTCTCACTGCTCATGATGATGATGATGATGATGAT 2040
Qy 2095 CAGATTGGGGAATGAAGTTGGACTACAGAGTTCATCTGGATATGATGATGATGATGATGAT 2154
Db 2041 TCGCTTTGGACAGCTTTTACACAGAAAGCGCTCTCCAGGCTTTGACCGTTCGCGAA 2100
Qy 2155 CGCGTCTATTTGGGAATGAAGCTTTCGATCTGACCTACCTGAGGAGGCTTACACACAGA 2214
Db 2101 TGCTGAGATTGGGAATGAAGCTTTCGATGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2160
Qy 2215 ACCTGCTGCTTTCGCTATCTATAGGTTGAAGAGCGCGCATGA 2256
Db 2161 ACATTGGCTGCTCAGGATATACAGGTTAAAGGACCTGGATAA 2202

RESULT 11

ADD94791
ID ADD94791 standard; DNA; 3094 BP.

XX AC ADD94791;

XX XX 29-JAN-2004 (first entry)

XX DT

XX DE Mouse ITM1 gene sequence.

XX XX

KW source of immunodominant MHC-associated peptide; SIMP; MHC;

KW major histocompatibility complex; human leukocyte antigen; HLA;

KW cystostatic; immunosuppressive; antineoplastic; gene therapy; cancer;

KW lung cancer; intestine cancer; sarcoma; prostate cancer;

KW testicular cancer; breast cancer; melanoma; pancreatic cancer;

KW haematological cancer; immune response; lymphoid cell proliferation;

KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;

XX murine; gene; ds; ITM1.

XX OS Mus musculus.

XX XX

XX PN W02003054008-A2.

XX XX

XX PD 03-JUL-2003.

XX XX

XX PD 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.
 PR (COMP-) COMPATIGENE INC.
 PA
 XX
 PI Perreault C, McBride K;
 DR WPI; 2003-559122/52.
 XX P-PSDB; ADD94792.
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Disclosure; SEQ ID NO 9; 66pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the mouse ITM1 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 3094 BP; 758 A; 651 C; 703 G; 982 T; 0 U; 0 Other;
 Query Match 26.1%; Score 631.2; DB 10; Length 3094;
 Best Local Similarity 58.2%; Pred. No. 2.1e-156;
 Matches 1246; Conservative 0; Mismatches 803; Indels 93; Gaps 4;
 118 CAGCCTATACCTTCGCCATCTGCTATATCGCTGCTGCGCGGATTTCTCTCGCCT 177
 156 CACACTCTAAAGCTTCTCATCTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 215
 178 CTTCGCCGTCATCCGCTTCGAGTCGATATCCATGAGTTGATCGCTGCTCAACTACCG 237
 216 TTTTGTCTGCTGAGATTTGAAAGTGTCTATCCATGATGTTGATCGTACTTTAATTATCG 275
 238 GGCACCCGCTACATGCTGCGAGATGTTGGTACACTTCTCACTGCTTCCGACGCG 297
 276 GACTACCCGGTTCTGCTGAGGAGGGTTTATAAATTCATAACTGCTTGTGATGACCG 335
 298 CGCATGTATCCGCTCGGCGAGATTTGGGCGGTACCGTCTATCCGCGCTGATGATAC 357
 336 GGCTTGTATCCCTTTGGCGCGAATCATTTGAGGAACTTTACCAGGTTTAATGATCAC 395
 358 GTCCGCGGGAATCCATTTGGCTGCTGCGACGCTACATACACACCGTCCATATTCGATAC 417
 396 TTTCTGCTGCAATACATGATCTCCATTTCTTCCATATCATATGATGATGCGAATGT 455
 418 CTGCGTGTCTGCGCGCGATTTCTAGTGGCGCTGACCTCCATCTCCACTACCTGCTGAC 477
 456 CTGTGTTTCTGCGCGCCACTTTTCTCTCTTTTCCACCACTGTTAGTACCACTTAC 515
 478 CAAGAGCTGTGTCGCGCGCGCGCGCTCTTCGCGCGCGCTTCATCGCATCGCTGCC 537
 516 CAAAGAGCTCAAGAGATGAGAGCTGGCTCTTCTGCTGCGCATGATGCTGTAGTGTCC 575
 538 TGGCTACATCAGTAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 597
 576 TGGGTATATTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 635

QY 598 CCTGCACTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 657
 DB 636 CATGCTGCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTT 695
 QY 658 GGCGGAGCGCTTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 717
 DB 696 TGCAAGTGTGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 755
 QY 718 CATCACTGATACCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
 DB 756 GATCACTTGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 815
 QY 778 TCTGCTGACAGCTACAGCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 837
 DB 816 GATCTACGTAGCTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 875
 QY 838 CTTCTGCTGATTCCTCAACGATACGACAGTGAACATGCTGCTGCTGCTGCTGCTGCTGCT 897
 DB 876 CTTTGTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 935
 QY 898 GCTCTTATGCGCGCTGCGCACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 957
 DB 936 TCTCTGCTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 995
 QY 958 CCGGAGCTGTTCTGCTGCGCGGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
 DB 996 CGAAGTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1055
 QY 1018 GGTGCTCACTGCTGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1077
 DB 1056 TCTCTCTATGCTTAACAGGAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1115
 QY 1078 TACTGCTGACGACAGATCCCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1137
 DB 1116 TCTCTTATGTAAGAATAACATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1175
 QY 1138 CACTTGTGTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1197
 DB 1176 AACCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1235
 QY 1198 GTGCTACTGCTACAGCAGATCAACGACGCGCTTCTTCTTCTTCTTCTTCTTCTTCTT 1257
 DB 1236 CTATCTACTCTTACACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1295
 QY 1258 TCGGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
 DB 1296 CAGCATGTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1355
 QY 1318 CATGCTGCGCGAGTGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1377
 DB 1356 CATCTTCTTCTGCAATGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1406
 QY 1378 TAAGCGAATGGSCACAGCCATTAAGCGCACCCCAAGTGAAGTGAAGTGAAGTGAAGT 1437
 DB 1407 ----- 1406
 QY 1438 TGAGAAGAGAGCGCTGTACGACAGGCTGCGCAAGCTGAAGCATCGTACTAAGCATGATGC 1497
 DB 1407 TCTGGACATAGTCTCCGACAGACAGAGACAGACAGACAGAGTCTTCTTCTTCTTCTTCT 1466
 QY 1498 CCAGCAGGATCTGCGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1557
 DB 1467 TAAGAATGAGG-----TGGCAGTGGGATGATGATGATGATGATGATGATGATGATG 1508
 QY 1558 GCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1617
 DB 1509 TTTTCTCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1568
 QY 1618 CTCCTATGCTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1677
 DB 1569 CTCCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1625

1678 CAGAGAGGCTTACTGCTTTCGAGAACACATCGCGATGATGCTCGCGTTATGCTTG 1737
 1626 CCGAGAGAGCTTATTTGGTCCGTCACATATCTCCAGAGATGAAAAGTCAATGTCATG 1685
 1738 GTGGGATACGATACCGATAGCGGGAATGCAAAAGACGAGCGTAGTGGATATAA 1797
 1686 GTGGGATATGCTTACCAATTTACTGCAATGCAAAATCGGCAATTTTAGTGACATAA 1745
 1798 TAGCTGGAACAATAGTCAATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGAA 1857
 1746 CACATGGATATATACCATATTTCTCGATAGGAGCAATGATGATGCAAGAGAAA 1805
 1858 GTCTACGAAATATGACATCTTTGACGTGGAATGCTTCAACCGAGGAGAA 1857
 1806 AGCCTATGAAATCATGAGGAGCTTGTATGATGCTATGCTATTTTGGAGGCT 1865
 1918 GATCGGCTATCTGGCGATGATATCAACAAGTTCTGTGGATGTCGGAATGCTGAGG 1977
 1866 TACTGGTATCTTCGATGATATCAACAAGTTCTGTGGATGTCGGAATGAGGAG 1925
 1978 ---AGAGCATCCAGGACATTAAGAAAGGATTAATTTTACCAGCGCGGTGAATTCAG 2034
 1926 CACAGAGACAGGAGACATTAAGGAGAGATGACTTACTATCTCTACTGCGGAATTCG 1985
 2035 GGTAGATCGGAGAGTCTCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2094
 1986 TGTGATCGTGAAGGTTCTCGGCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2045
 2095 CAGATTCGGGAAATGAAGTTGGAATACAGAGGTCATCTGATATGATGACACAGTAA 2154
 2046 CCGCTTTGGGAGGCTACACAGAGCAAGCGTCCACAGGCTTTGACCGTGTTCGAAA 2105
 2155 CCGCTGCAATGGGAATAGGACTTCGATCTGACCTACCTGAGAGAGGCTTACACACAGA 2214
 2106 TCGTGAATGTTGATTAAGACTTTGAGCTTGATGCTCTGAGAGAGGCTATACACAGA 2165
 2215 ACATGCTGTTGTCGATCTATAGGTTGAAGCGGATGA 2256
 2166 ACATGCTGCTAGTCAGATATACAGGTAAGGACCTGGATAA 2207

RESULT 12
 ADB69900
 ID ADB69900 standard; DNA; 2466 BP.
 AC ADB69900;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE C. neoformans open reading frame SEQ ID NO:2305.
 XX
 KW ds; gene; fungicide; gene therapy; infection.
 XX
 OS Cryptococcus neoformans.
 XX
 PN WO2003052076-A2.
 XX
 PD 26-JUN-2003.
 XX
 PF 17-DEC-2002; 2002WO-US040225.
 XX
 PR 17-DEC-2001; 2001US-0341261P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Zamudio C, Eroshkin AM;
 XX
 PI WPI; 2003-533017/50.
 DR P-PSDB; ADB70261.
 XX
 PT New nucleic acid, useful for preparing a composition for treating an
 XX infection caused by Cryptococcus neoformans.

PS Claim 2; SEQ ID NO 2305; 136pp; English.
 XX
 CC The invention relates to a novel purified or isolated Cryptococcus
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;
 Query Match 25.7%; Score 620.2; DB 10; Length 2466;
 Best Local Similarity 56.7%; Pred. No. 1.5e-153;
 Matches 1210; Conservative 0; Mismatches 908; Indels 15; Gaps 3;
 QY 119 AGCTAATCACTTCGCGATCTCTGTAATCGCTGGCTGGCGGATTTTCTCTCGCCTC 178
 DB 226 AGCTTGTTCGCTTCATCATCTCGCGTTAATATGCGGTGCTGCGATTTGAAGTCGATTG 285
 QY 179 TTGCGCGTCATCCGTTTCGAGTCGATTAATCATGAGTTTTCATCGGTGTTCAACTACCGG 238
 DB 286 TTGCGGTGATCAGATTCGAATCTGTCTATCCAGATTTGACCCCTGTTCAACTACCGA 345
 QY 239 GCCACCGCTACATGTTGTCAGATGTTGGTACAACTTCTCTCAACTGTTTCGACGAGCGC 298
 DB 346 GCTCTGAAGTCTTGTGTTAACAAGGTTTCTAGAGTTCTGGAATCTGTTTGAACCTCC 405
 QY 299 GCATGATATCCGTCGCGAGGATTTGGGCGGTACCGTCTATCCGCGCTGATGATTACG 358
 DB 406 GCTTGTGTTACCTTCGCGAGACATGTCGTACCAAGCTCTATCTGCGTTGATGTCAG 465
 QY 359 TCCGCGGAATCCATTTGGCTCTGACGATCTAACATACCGGTTCATATCGTGATC 418
 DB 466 TCTGACTGATTTGGCATGCTCTTCGCGCAATCAATATGCCCGTGGACATTCGCAATGC 525
 QY 419 TCGGTGTTCTTCGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACC 478
 DB 526 TGTGCTCTCTTGACCTGGAATTTCTGGAATGACTGCTGGGGGACTTATCTTTTCAAC 585
 QY 479 AAGGAGCTGTGTCGCGCGCGCGCTCTTCGCGCGCAGCTTTCATCGCCTGCTGCT 538
 DB 586 ACTGAATGTCTACACCATCAGCTGCTTATTTGCGCGCGCTTTTCAATTTGACCTT 645
 QY 539 GGTACATCAGTAGGTGCTGCTGATCTGATACGATAACGAGGCGATTGGCATATTTCGCC 598
 DB 646 GGATACATCTCTGATCTGTCGCGGTTCTTATGACAAAGCAATTTGCACTTTCTTC 705
 QY 599 CTGCAATTCACCTACTCTCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
 DB 706 TTGATGAGCTCTTCTACTCTTGTGATTAAGCGCGTCAAAACCGGTAGCTCATTTGGGGT 765
 QY 659 GCGCGAGCGCTTTGTCCTACTTCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
 DB 766 ATGATCATCTGCTTGTCTACGCGTGGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 825
 QY 719 ATCAACCTGATACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778
 DB 826 ACCAATCATGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
 QY 779 CTGCTGACGCTACAGCACTTCTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 838
 DB 886 CTTTATACGCTTACTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945
 QY 839 TTGCTGGATTCACACGATACGCAACGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
 DB 946 TTGCTGGATTCCTCCCGCATCCGAACTCTGAGCAGATGCGCGCTGCTGCTGCTGCTGCT 1005
 QY 899 CTCTTATGCGCGTGGCGCACTTTCGCGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958

| | | | | | |
|-----------------------|------|--|---------------------|-----------------|--------------------|
| Query Match | | 25.5% | Score 615.6; | DB 5; | Length 2760; |
| Best Local Similarity | | 58.3% | Pred. No. 2.7e-152; | | |
| Matches 1250; | | Conservative | 0; | Mismatches 799; | Indels 95; Gaps 6; |
| QY | 118 | CGACCTAATCAGCTTCGGCATCTCTGCTAATCGCTGGCTGGCCGGATTTTCCTCTGSCCT | 177 | | |
| DB | 149 | CACACTTTTGAAGCTTCTCATCTGTCATAGGCTGCTGTATATTCCTTCTCCACTCGTCT | 208 | | |
| QY | 178 | -CTTCCGCGTCACTCGTTTCAGTGCATTTATCCATGAGTTTGATCCGTTGATCCACTACC | 236 | | |
| DB | 209 | GTTTTGTCTGCTGAGATTGAAGTGTTATCCATGAGTTTGATCCGTTGATCCACTACC | 268 | | |
| QY | 237 | GGCCACCCGCTACATGCTGTCAGATGTTGGTGAACAATCTTCCCTCACTGTTTCGACAGC | 296 | | |
| DB | 269 | GGACTACCAAGTTCCTGCTGAGGAGGGGTTTAAATTCATTAACCTGTTTGAATGACC | 328 | | |
| QY | 297 | GGGCTATGATCCGCTCGGCAGGATTGTGGCGGTACCGTCTATCCCGGCTCATGATTA | 356 | | |
| DB | 329 | GAGCTGTGTAACCTTTGGGAGCAATCATTTGGAGGAACAATTTACCCAGGTTTATGATCA | 388 | | |
| QY | 357 | CGTCCGGGGATTCATTTGGCTGCTGACGTA-CTCAACATACCGGTCCATATCGTGAC | 415 | | |
| DB | 389 | CTCTGCTGCAATCAACATGATCTCCATTTTTCACATCAACATCGATCGAATCGGAAT | 448 | | |
| QY | 416 | ATCTGCTGTTCTCTGGCGCCGATCTTCACTGCTGACCTCCATCTCCACTACCTGCTG | 475 | | |
| DB | 449 | GTCTGTGTCTCTGGCCCTCTCTTCTCTCTCTTCCACCAATCGTCACTACCACTT | 508 | | |
| QY | 476 | ACCAAGGAGCTGTGTCTCGGGCGCCGCTCTTTCGGCGCCAGCTTCATCGCCATCGTG | 535 | | |
| DB | 509 | ACCAAGAGCTCAAGATGCAAGGGCTGGGCTTCTGCTGCTGCATGATGCTGTAGTT | 568 | | |
| QY | 536 | CTGCTACATCAGTAGTGTGGTGGCTGATGATGATGATGATGATGATGATGATGATGATG | 595 | | |
| DB | 569 | CTGATATATCTCCGATCTGTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 628 | | |
| QY | 596 | GCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 655 | | |
| DB | 629 | TGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 688 | | |
| QY | 656 | TGCGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 715 | | |
| DB | 689 | GCAGCTAAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 748 | | |
| QY | 716 | ATCATCACTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 775 | | |
| DB | 749 | CTGATCAATTAATCTCTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 808 | | |
| QY | 776 | CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 835 | | |
| DB | 809 | CGGATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 868 | | |
| QY | 836 | CCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 895 | | |
| DB | 869 | TCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 928 | | |
| QY | 896 | GTGCTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 955 | | |
| DB | 929 | GTGCTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 988 | | |
| QY | 956 | TTCCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1015 | | |
| DB | 989 | TTTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1048 | | |
| QY | 1016 | GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1075 | | |
| DB | 1049 | GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1108 | | |
| QY | 1076 | GATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1135 | | |
| DB | 1109 | GATCTCTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1168 | | |
| QY | 1136 | ACCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1195 | | |

| | | | | | |
|----|------|--|------|--|--|
| DB | 1169 | ACACCTGCT | 1228 | | |
| QY | 1196 | GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1255 | | |
| DB | 1229 | CTCTATTAATCT | 1288 | | |
| QY | 1256 | AGTGGGTTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1315 | | |
| DB | 1289 | ACCAGCATGATCT | 1348 | | |
| QY | 1316 | TGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1375 | | |
| DB | 1349 | TGCATCT | 1384 | | |
| QY | 1376 | TCTAAGCAATGGGCGACACCATTAAGCGACACCAAGTGAAGTGAAGTGAAGTGAAGTGA | 1435 | | |
| DB | 1385 | -----TCC | 1387 | | |
| QY | 1436 | ATTGAGAAAGACGCTGTACGACAGCTGGGAGCTGAAGCATCGTACTAAGCATGAT | 1495 | | |
| DB | 1388 | ACATACATGAAGAATCTGGACATTAAGTCTCCAGACAAGAAAGAGCAAGCAACAGGAT | 1447 | | |
| QY | 1496 | CCCCAGCAGGATACCTGGCGTCAGCTCCCACTGAAGATGATTTTGGCGCTTCTA | 1555 | | |
| DB | 1448 | TCCACTACCTA-----TTAAGATGAAGTGGCAAGTGGGATGATCTGGTCTGGCT | 1501 | | |
| QY | 1556 | ATGCTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1615 | | |
| DB | 1502 | TCTTCTCTCATCAGCTACCTTTCTTCACTGGGTGACCACTGAGGCTGCTCTCTCTCT | 1561 | | |
| QY | 1616 | CCCTCCATGCTTGGGCTTCCCAACAGTCAAGATGGATCCCGCAACATTTTGAAGCAT | 1675 | | |
| DB | 1562 | CGTCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1618 | | |
| QY | 1676 | TTCAGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1735 | | |
| DB | 1619 | TTCAGAGGATGATTAATTTGGCTTCTGCTAATACTCCAGAGGATGCAAGGCTGCTGCTG | 1678 | | |
| QY | 1736 | TGGTGGATACCGATACCGGATAGCGGAAATGGCAACAGACAGGCTAGTGGGATAT | 1795 | | |
| DB | 1679 | TGGTGGATATGCTATCAGATTTACAGCTATGGCAACCGAACAATTTTGTGGACAAT | 1738 | | |
| QY | 1796 | AATAGTGGAAATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1855 | | |
| DB | 1739 | AACACATGGAATAATCCCATATTTCTGAGTAGGCGAGCAATGGCTCCACAGAGGAA | 1798 | | |
| QY | 1856 | AAGTCTTACGAAATTTATGACATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1915 | | |
| DB | 1799 | AAAGCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTGCTGCTGCTGCTGCTG | 1858 | | |
| QY | 1916 | GTGATCGCTATTTCTGGGATGATATCAACAAGTTCCTGCTGCTGCTGCTGCTGCTGCTG | 1975 | | |
| DB | 1859 | CTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 1918 | | |
| QY | 1976 | GG---AGAGCATCCCAAGGACATTAAGGAAAGGATTTTACCGACCGGCTGAATTC | 2032 | | |
| DB | 1919 | AGCAGATACAGGCAACATATCAAGGAGATGACTATATATCTCAACTGGGAGTTTC | 1978 | | |
| QY | 2033 | AGGTAGATCCGAGGATGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 2092 | | |
| DB | 1979 | CGTGGACCGTGAAGGTTCTCAGTGTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 2038 | | |
| QY | 2093 | TACAGATTCGGGGAATGAAGTTGGACTACAGAGGCTCCATCTGATATGATCCACACGT | 2152 | | |
| DB | 2039 | TATCGTTTGGACAGTTTACACAGAGCCAGGCTCTCCAGCTTTTGAACGCTGCTCGA | 2098 | | |
| QY | 2153 | AACGCGCTCATTTGGGATTAAGGATTCGATCTGACCTACCTGAGGAGGCTTACACACA | 2212 | | |
| DB | 2099 | AATGCTGAGATTTGGGAAATAAGACTTTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTG | 2158 | | |
| QY | 2213 | GAAACACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 2256 | | |

| | | | | |
|-----------------------|---|---|------|--|
| Db | 2159 | GAACATTGGCTGGTTCAGGATATACAAAGGTAAGGACCTGGATAA | 2202 | Matches 1071; Conservative 0; Mismatches 655; Indels 66; Gaps 5; |
| RESULT 14 | | | | |
| ABT18969 | | | | |
| ID | ABT18969 | standard; DNA; 1848 BP. | | |
| AC | ABT18969; | | | |
| XX | | | | |
| DT | 16-APR-2003 | (first entry) | | |
| XX | | | | |
| DE | Aspergillus fumigatus essential gene #1327. | | | |
| XX | | | | |
| KW | Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection; | | | |
| XX | cancer; contamination; biofilm; antibody; immune response; ds. | | | |
| OS | Aspergillus fumigatus. | | | |
| XX | | | | |
| PN | WO200286090-A2. | | | |
| XX | | | | |
| PD | 31-OCT-2002. | | | |
| XX | | | | |
| PF | 23-APR-2002; 2002WO-US013142. | | | |
| XX | | | | |
| PR | 23-APR-2001; 2001US-0285697P. | | | |
| XX | | | | |
| PR | 27-JUN-2001; 2001US-0287066P. | | | |
| XX | | | | |
| PR | 05-JUN-2001; 2001US-0295890P. | | | |
| XX | | | | |
| PR | 09-JUL-2001; 2001US-0303899P. | | | |
| XX | | | | |
| PR | 31-AUG-2001; 2001US-0316362P. | | | |
| XX | | | | |
| PA | (ELIT-) ELITRA PHARM INC. | | | |
| XX | | | | |
| PI | Jiang B, Tishkoff D, Zamudio C, Broshkin AM, Hu W, Lemieux SM; | | | |
| XX | | | | |
| DR | WPI; 2003-093124/08. | | | |
| XX | | | | |
| PT | New purified or isolated nucleic acids of essential genes of Aspergillus | | | |
| XX | fumigatus, useful for treating or preventing infections by A. fumigatus, | | | |
| PT | or for treating a non-infectious disease in a subject e.g. cancer. | | | |
| XX | | | | |
| PS | Disclosure; Page; 175pp; English. | | | |
| XX | | | | |
| CC | The invention relates to novel purified or isolated nucleic acids of | | | |
| CC | essential genes of Aspergillus fumigatus. The isolated nucleic acids of | | | |
| CC | the invention are used to treat or prevent infections by a pathogenic | | | |
| CC | organism such as A. fumigatus, to treat a non-infectious disease in a | | | |
| CC | subject (e.g. cancer), to prevent or contain contamination of an object | | | |
| CC | by A. fumigatus, or to prevent or inhibit formation on a surface of a | | | |
| CC | biofilm comprising A. fumigatus. The polynucleotides are useful for | | | |
| CC | expressing recombinant protein for characterisation, screening or | | | |
| CC | therapeutic use, as markers for host tissues in which the pathogenic | | | |
| CC | organisms invade or reside, for comparing with the DNA sequence of A. | | | |
| CC | fumigatus to identify duplicated genes or paralogues having the same or | | | |
| CC | similar biochemical activity and/or function, for comparing with DNA | | | |
| CC | sequences of other related or distant pathogenic organisms to identify | | | |
| CC | potential orthologous essential or virulence genes, for selecting and | | | |
| CC | making oligomers for attachment to a nucleic acid array for examination | | | |
| CC | of expression patterns, for raising anti-protein antibodies, as an | | | |
| CC | antigen to raise anti-DNA antibodies or to elicit another immune | | | |
| CC | response, and for identifying polynucleotides encoding the other protein | | | |
| CC | with which binding occurs or to identify inhibitors of the binding | | | |
| CC | interaction. The polypeptides may be used to raise antibodies or to | | | |
| CC | elicit immune response, as a reagent in assays designed to quantitatively | | | |
| CC | determine levels of the protein in biological fluids, as a marker for | | | |
| CC | host tissues in which pathogenic organism invade or reside, and to | | | |
| CC | isolate correlative receptors or ligands in the case of virulence | | | |
| CC | factors. This polynucleotide sequence represents one of the essential | | | |
| CC | genes of Aspergillus fumigatus of the invention | | | |
| XX | | | | |
| SQ | Sequence 1848 BP; 363 A; 510 C; 447 G; 528 T; 0 U; 0 Other; | | | |
| Query Match | 23.3%; Score 562; DS 8; Length 1848; | | | |
| Best Local Similarity | 59.8%; Pred. No. 3.6e-138; | | | |

QY 1377 CTAAGCAATGGGACAGCCATAAGCGCAGCCAGGAGTGGATGAGCTGAGGATCCA 1436
 Db 1129 -----GGGACTACCTCCCGACACCGAGCGTCTGAAGC----- 1160
 QY 1437 TTGAGAAGAGACGGCTGTAGCAGACGGCTGGCAGAGCTGAGCAGCTGTAAGCATGATG 1496
 Db 1161 -----GAAACGAAATGAAGACTCGTCTTCAACACCTCTTGGTCTGATTC 1204
 QY 1497 CCGAGCAGATGACTGGCGTCAAGCTCCAACT---GAAGAGTATTGTTATTTGGCCGCTTC 1553
 Db 1205 GGAAGCCCAATGTTGGAATCACTCCCATGTTCTTAAGATTATAGTGACGGCGTCTGTTG 1264
 QY 1554 TAACTGTTGATGATGTTCTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613
 Db 1265 TCGTCTACCTGCTGCTGTTGTTGGCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1324
 QY 1614 GTCCCTCCATGCTTGGCTTTTCCACACAGTCAAGATGATGATGATGATGATGATGATGATG 1673
 Db 1325 CTCCTTCCTGCTTGGCTAGTGGATGCT---GACGGAAGCCAATACATCATTTGACG 1381
 QY 1674 ATTTGAGAGAGCTTACTGCTTTCGAGACACTGCGGATGATGCTCGGTTATGT 1733
 Db 1382 ATATCGTGAAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1441
 QY 1734 CTGTGTGGGATTACGATACAGATAGCGGAATGGCAAAACAGAACGACGCTAGTGGATA 1793
 Db 1442 CATGTGGGATTATGGGTATCAATCGGTGGCATGGCGGCGCCCAACCTTGGTTGACA 1501
 QY 1794 ATAACTGTGAGACATAGTACATAGGCTGGTGGCAGGCAATGCTTCAACCGAGG 1853
 Db 1502 ACAACACCTTGAACACCAACCATATGTACGGTGTGTAAGGCGATGAGTCAACGAGG 1561
 QY 1854 AGAAGTCTTACGAAATATGACATCTTGTGACGTGGACTAGTGTGCTGATCTTTGGCG 1913
 Db 1562 AAGTCAGCTACCCATCTCCGCGCAGCATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 1621
 QY 1914 GTGTATGCTGCTTCTGCGGATGATATCAACAGATTCCTGATGATGCTGATGCTGCTGCTG 1973
 Db 1622 GTCTGCTAGTATTCTGCGGATGATATCAAAATCTTATGATGCTGCTGCTGCTGCTGCTGCT 1681
 QY 1974 AGGAGAGCATCCCAAGAGCATTAAAGGAAGCGATTACTTTACCGACCGCGGTGAATCA 2033
 Db 1682 AAGGTATCTGGCCCGATGAGGTAAAGAGCGGACTTCTTATGTCAGCGGTGAATATC 1741
 QY 2034 GGGTAGATCCCAAGGTGCTCGGCGCTGCTCAACTGCTTATGATCAAAAT 2085
 Db 1742 GTGTGAGATGAGCGACCCCAACTATGCGCAACAGCTTATGATGATGATGATGATGATGAT 1793

RESULT 15
 ID ADD94787
 XX AC ADD94787 standard; DNA; 2733 BP.
 XX AC ADD94787;
 XX DT 29-JAN-2004 (first entry)
 XX DE Yeast STT3 gene sequence.
 XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestinal cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; yeast;
 KW gene; ds; STT3.
 OS Saccharomyces cerevisiae.
 XX PN W02003054008-A2.
 XX

PD 03-JUL-2003.
 XX 18-DEC-2002; 2002WO-CA001967.
 XX 20-DEC-2001; 2001US-00028384.
 XX (COMP-) COMPATIGENE INC.
 XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 XX P-PSDB; ADD94788.
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancer, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX Disclosure; SEQ ID NO 5; 66pp; English.
 XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancer, for example lung cancer,
 CC intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response.
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the yeast STT3 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from Genbank.
 XX
 SQ Sequence 2733 BP; 712 A; 597 C; 583 G; 841 T; 0 U; 0 Other;
 Query Match 21.6%; Score 521.8; DB 10; Length 2733;
 Best Local Similarity 54.8%; Pred. No. 2e-127;
 Matches 1211; Conservative 0; Mismatches 877; Indels 123; Gaps 4;
 QY 105 TGGCTGGCTACAGCAGCTTAATCACTTCGCCATCTGCTAATCGGCTGGCTGGCGGAT 164
 Db 244 TGCTGTGTTTCAGACCATCTCAAGCTCGTCTATCTTCGTGGCGATTTTGGGGCTGCCA 303
 QY 165 TTTCTCTCGGCTCTTGGCGCTCATCGGTTTCGAGTCGATTATCCATGATGATTTGATCCGT 224
 Db 304 TATCATCAGTTTGTTCGAGTCATCAAAATTTGAGTCTATTATCCATGAATTCGACCCCT 363
 QY 225 GGTCAACTACGGGCGCACCGCTACATGTGAGATGTTGTTGATCACTTCTCAACT 284
 Db 364 GGTCAATATATAGGGCTACCAATATCTCGTCAACAAATTCGTTTACAAAGTTTGAAT 423
 QY 285 GGTTCGACGAGCGGCATGTTATCCGCTCGCGAGGATTTGGCGGTACCGTCTATCCCG 344
 Db 424 GGTTCGACGAGCGGTACCTGTTACCCCTCGAGGGTTACTGGAGGACTTTATATCTG 483
 QY 345 GCCTGATGATGATGCTCGCGGGAATC---CATTTGGCTGCTGACGACTCAATACCGG 401
 Db 484 GTTGTATGACACTAGTGTGCTGTTCTGTCGCAACCGCTTCGCAACTGGTTGGGTTGCCA 543
 QY 402 TCCATATTCGTGACATCTGCTGTGTTCTCGCGGCGGATCTTCACTGGCTGACCTCCATCT 461
 Db 544 TTGACATCAGAAACGTTTGTGTCTATTTCGCGCCACTATTTCTGGGGTCAACCGCTGG 603
 QY 462 CCACTACTGCTGACCAAGAGCTGTGTCGCGGGCGCGGCTCTTCGCGCCAGCT 521
 Db 604 CGACTTACGAATTTACGAAAGAGATTAAAGATCGACGCTGGCTTTTGGCTGCTGGTT 663

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:40:24 ; Search time 192.945 Seconds
(without alignments)
8903.949 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctaagcgaagatgtgtcg.....ccaaaaaataaaaaaa 2417

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 889 | 36.8 | 900 | 4 | US-09-270-767-12331 |
| 2 | 756.4 | 31.3 | 1660 | 4 | US-09-270-767-158 |
| 3 | 756.4 | 31.3 | 1660 | 4 | US-09-270-767-15440 |
| C 4 | 643.2 | 26.6 | 1660 | 4 | US-09-270-767-158 |
| C 5 | 643.2 | 26.6 | 1660 | 4 | US-09-270-767-15440 |
| C 6 | 552.4 | 22.9 | 560 | 4 | US-09-270-767-1318 |
| C 7 | 552.4 | 22.9 | 560 | 4 | US-09-270-767-16600 |
| 8 | 523.4 | 21.7 | 2157 | 4 | US-09-614-221A-318 |
| 9 | 349.2 | 14.4 | 1386 | 4 | US-09-270-767-11648 |
| 10 | 284.4 | 11.8 | 914 | 4 | US-09-270-767-12856 |
| 11 | 279.6 | 11.6 | 507 | 4 | US-09-513-999C-1965 |
| 12 | 244.4 | 10.1 | 867 | 4 | US-09-248-796A-3089 |
| 13 | 173 | 7.2 | 616 | 3 | US-09-328-111-332 |
| 14 | 171.2 | 7.1 | 503 | 4 | US-09-270-767-27262 |
| 15 | 150.2 | 6.2 | 487 | 3 | US-09-385-982-213 |
| 16 | 134 | 5.5 | 307 | 4 | US-09-702-705-1271 |
| 17 | 134 | 5.5 | 307 | 4 | US-09-736-457-1271 |
| 18 | 134 | 5.5 | 307 | 4 | US-09-614-124B-1271 |
| 19 | 134 | 5.5 | 307 | 4 | US-09-671-325-1271 |
| 20 | 134 | 5.5 | 307 | 4 | US-09-658-824-1271 |
| 21 | 133.6 | 5.5 | 487 | 4 | US-09-702-705-1655 |
| 22 | 133.6 | 5.5 | 487 | 4 | US-09-736-457-1655 |
| 23 | 133.6 | 5.5 | 487 | 4 | US-09-614-124B-1655 |
| 24 | 133.6 | 5.5 | 487 | 4 | US-09-671-325-1655 |
| 25 | 133.6 | 5.5 | 487 | 4 | US-09-658-824-1655 |
| 26 | 133 | 5.5 | 133 | 4 | US-09-270-767-28082 |
| 27 | 132.4 | 5.5 | 307 | 4 | US-09-702-705-375 |

28 132.4 5.5 307 4 US-09-702-705-588 Sequence 588, App
29 132.4 5.5 307 4 US-09-736-457-375 Sequence 375, App
30 132.4 5.5 307 4 US-09-736-457-588 Sequence 588, App
31 132.4 5.5 307 4 US-09-614-124B-375 Sequence 375, App
32 132.4 5.5 307 4 US-09-614-124B-588 Sequence 588, App
33 132.4 5.5 307 4 US-09-671-325-375 Sequence 375, App
34 132.4 5.5 307 4 US-09-671-325-588 Sequence 588, App
35 132.4 5.5 307 4 US-09-589-184-375 Sequence 375, App
36 132.4 5.5 307 4 US-09-589-184-588 Sequence 588, App
37 132.4 5.5 307 4 US-09-658-824-375 Sequence 375, App
38 132.4 5.5 307 4 US-09-658-824-588 Sequence 588, App
39 131.4 5.4 594 4 US-09-248-796A-3090 Sequence 3090, App
40 125 5.2 292 4 US-09-313-294A-4442 Sequence 4442, App
41 115.4 4.8 245 4 US-09-513-999C-21090 Sequence 21090, A
42 102 4.8 288 4 US-09-313-294A-6705 Sequence 6705, App
43 97.8 4.0 291 4 US-09-313-294A-4834 Sequence 4834, App
44 96.6 4.0 414 4 US-09-513-999C-271 Sequence 271, App
45 93.2 3.9 302 4 US-09-313-294A-6869 Sequence 6869, App

ALIGNMENTS

RESULT 1

US-09-270-767-12331

; Sequence 12331, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09270767

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 12331

; LENGTH: 900

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-12331

Query Match 36.8%; Score 889; DB 4; Length 900;
Best Local Similarity 100.0%; Pred. No. 4e-239;
Matches 889; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAAGCGAAGATGTGCTGCTTCAGATTCAGATCGGTTATATTTTCGAGTTACTGCGTG 60
Db 12 TCTAAGCGAAGATGTGCTGCTTCAGATTCAGATCGGTTATATTTTCGAGTTACTGCGTG 71
QY 61 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
Db 72 GAATTTGGGACATGAATCGGACCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 131
QY 121 CCTAATCACTTCGCCATTCCTGCTTAATCGCTGGCTGGCGGATTTTCCTCTCCCTCTT 180
Db 132 CCTAATCACTTCGCCATTCCTGCTTAATCGCTGGCTGGCGGATTTTCCTCTCCCTCTT 191
QY 181 CGCGTCATCCGTTTCGAGTCGATTATCCATGATTTGATCCGTCGGTGGTCAACTACCGGGC 240
Db 192 CGCGTCATCCGTTTCGAGTCGATTATCCATGATTTGATCCGTCGGTGGTCAACTACCGGGC 251
QY 241 CACCGCTPACATGTGTCAGAAATGGTTGGTCAACTTCCTCAATGGTTCGACGAGCGGC 300
Db 252 CACCGCTPACATGTGTCAGAAATGGTTGGTCAACTTCCTCAATGGTTCGACGAGCGGC 311
QY 301 ATGGTATCCGTCGGCAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTC 360
Db 312 ATGGTATCCGTCGGCAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTC 371
QY 361 CGCGGGAATCCATTTGGCTGTGTCACGCTACACATACCGGTCCATTTTCGATGATCTG 420
Db 372 CGCGGGAATCCATTTGGCTGTGTCACGCTACACATACCGGTCCATTTTCGATGATCTG 431

| | | | | |
|-----|----|-------------------------------------|-----------------------------|-----|
| 421 | QY | CGTGTCTCGGCGCGGATCTTCAGTGGCGCTGA | CGTCCATCTCAACCTACCTCGTGACCA | 480 |
| | | | | |
| 432 | Db | CGTGTCTCGGCGCGGATCTTCAGTGGCGCTGA | CTCCACACTACCTCGTGACCA | 491 |
| | | | | |
| 481 | QY | GGAGCTGTGCTCCGCGGCGCGGCTCTTCGGCGG | CAGCTTCATCGCCATCGTGCCTGG | 540 |
| | | | | |
| 492 | Db | GGAGCTGTGCTCCGCGGCGCGGCTCTTCGGCGG | CAGCTTCATCGCCATCGTGCCTGG | 551 |
| | | | | |
| 541 | QY | CTACATCAGTAGGTCGGTGGCTGGATCGTACGATA | ACGAGGCAATGCCCATTATCGCCCT | 600 |
| | | | | |
| 552 | Db | CTACATCAGTAGGTCGGTGGCTGGATCGTACGATA | ACGAGGCAATGCCCATTATCGCCCT | 611 |
| | | | | |
| 601 | QY | GCAGGTCACCTACTTCCTGTGGGTGGCGCTCAGT | CGACACTGGATCGTGCTCTGGTCGCGC | 660 |
| | | | | |
| 612 | Db | GCAGGTCACCTACTTCCTGTGGGTGGCGCTCAGT | CGACACTGGATCGTGCTCTGGTCGCGC | 671 |
| | | | | |
| 661 | QY | CGCAGCCGCTTTGTCTACTTCTACATGCTGTCCG | CTCGGCTGGCTAGCTGTTTCATCAT | 720 |
| | | | | |
| 672 | Db | CGCAGCCGCTTTGTCTACTTCTACATGCTGTCCG | CTCGGCTGGCTAGCTGTTTCATCAT | 731 |
| | | | | |
| 721 | QY | CAACCTGATACCCCTGCACGCTTCCTGCTACTGCT | CATTATGGGAGAGTACTCGCCGCGTCT | 780 |
| | | | | |
| 732 | Db | CAACCTGATACCCCTGCACGCTTCCTGCTACTGCT | CATTATGGGAGAGTACTCGCCGCGTCT | 791 |
| | | | | |
| 781 | QY | GCTGACCAAGCTACAGCACTTCTACATCTCTGGAC | TGCTGTTCTCCATGCAGATCCCTT | 840 |
| | | | | |
| 792 | Db | GCTGACCAAGCTACAGCACTTCTACATCTCTGGAC | TGCTGTTCTCCATGCAGATCCCTT | 851 |
| | | | | |
| 841 | QY | CGTGGGATTCCAACCGGATACGACCAAGTGAAACA | CATGGCTGGCTGGGA | 889 |
| | | | | |
| 852 | Db | CGTGGGATTCCAACCGGATACGACCAAGTGAAACA | CATGGCTGGCTGGGA | 900 |

RESULT 2

```

US-09-270-767-158
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and
; FILE REFERENCE: File Reference: 73266
; CURRENT APPLICATION NUMBER: US/09/270-767-158
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-158

```

| Query Match | 31.3% | Score 756.4; | DB 4; | Length 1660; |
|-----------------------|-----------------|---|-----------|--------------|
| Best Local Similarity | 99.9%; | Pred. No.8.7e-202; | | |
| Matches 757; | Conservative 0; | Mismatches 1; | Indels 0; | Gaps 0; |
| QY | 1655 | TCCGCGCAACATTTTAGACGATTTTCAGAGAGGCTTACTACTGGCTTTCCGAGAACACATGCC | 1714 | |
| DB | 768 | TCCGCGCAACATTTTAGACGATTTTCAGAGAGGCTTACTACTGGCTTTCCGAGAACATGCC | 847 | |
| QY | 1715 | GATGATGCTCGCGTTATGTCTTCGGTGGGATTTACGGATACCAGATAGCGGGGAATGCGAAC | 1774 | |
| DB | 848 | GATGATGCTCGCGTTATGTCTTCGGTGGGATTTACGGATACCAGATAGCGGGGAATGCGAAC | 907 | |
| QY | 1775 | AGAACGACGCTAGTGGGATAATAATACGTGGAAACAATAGTACATAGCCGCTGGTTGGCAAG | 1834 | |
| DB | 908 | AGAACGACGCTAGTGGGATAATAATACGTGGAAACAATAGTACATAGCCGCTGGTTGGCAAG | 967 | |
| QY | 1835 | GCAATGCTTCACCGGAGGAGTCCCTACGAAATATGACATCTCTTGACGTGGATC | 1894 | |
| DB | 968 | GCAATGCTTCACCGGAGGAGTCCCTACGAAATATGACATCTCTTGACGTGGATC | 1027 | |
| QY | 1895 | GTTTGGTGATCTTTGGCGTGTGATCGGCTATCTTCGGCATGATATCAACAAGTTCCCTG | 1954 | |

| | | | |
|------|----|---|------|
| 1028 | Db | GTTTTGGTGATCTTTGGCGGTGTATCGGCTATTCTGGCGATGATATCAACAAGTTCCCTG | 1087 |
| 1955 | Qy | TGATGTGTCGGAATTGCTTGAGGAGAGCATCCCAAGGACATTAAAGAAAGCGATTACTTTT | 2014 |
| 1088 | Db | TGGATGTGTCGGAATTGCTGAGGAGAGCATCCCAAGGACATTAAAGAAAGCGATTACTTTT | 1147 |
| 2015 | Qy | ACCAGACCGCGGTGAATTCAGGGTAGATCCGAAGGTGCTCGGCCCTGTCAACTGTCCTT | 2074 |
| 1148 | Db | ACCAGACCGCGGTGAATTCAGGGTAGATCCGAAGGTGCTCGGCCCTGTCAACTGTCCTT | 1207 |
| 2075 | Qy | ATGTACAAATTAAGCTACTACAGATTTCGGGGAATTGAAGTTGGACTACAGAGTCCCATCT | 2134 |
| 1208 | Db | ATGTACAAATTAAGCTACTACAGATTTCGGGGAATTGAAGTTGGACTACAGAGTCCCATCT | 1267 |
| 2135 | Qy | GGATATGATCGCACAGTAAACCGCTCAATTGGGAATAAGGACTTTCGATCTGACCTACCTG | 2194 |
| 1268 | Db | GGATATGATCGCACAGTAAACCGCTCAATTGGGAATAAGGACTTTCGATCTGACCTACCTG | 1327 |
| 2195 | Qy | GAGGAGCGCTACACCCAGAACACTGGCTTGTTCGCATCTATAGGTTGAAGAGCGCGAT | 2254 |
| 1328 | Db | GAGGAGCGCTACACCCAGAACACTGGCTTGTTCGCATCTATAGGTTGAAGAGCGCGAT | 1387 |
| 2255 | Qy | GAGTTCAATAGACCATCTGTAAGACCAAGAGAGAGACGATTCTCTCCAGCAAACTTCATT | 2314 |
| 1388 | Db | GAGTTCAATAGACCATCTGTAAGACCAAGAGAGAGACGATTCTCTCCAGCAAACTTCATT | 1447 |
| 2315 | Qy | TCGGAAGAAGAACTCTAAGCGTCGAAGGGCTACATAGCAACCGACCGGTGTGTTTAG | 2374 |
| 1448 | Db | TCGGAAGAAGAACTCTAAGCGTCGAAGGGCTACATAGCAACCGACCGGTGTGTTTAG | 1507 |
| 2375 | Qy | GGAAAAAGAACTTGAATAAACCACAAAAAATAAAAAA | 2412 |
| 1508 | Db | GGAAAAAGAACTTGAATAAACCACAAAAAATAAAAAA | 1545 |

RESULT 3

```

US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 63517
; SOFTWARE: PatentIniver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

```

| | Query Match | 31.3% | Score 756.4; | DB 4; | Length 1660; |
|----|-----------------------|--|---------------------|---------------|-------------------|
| | Best Local Similarity | 99.9%; | Pred. No. 8,7e-202; | | |
| | Matches : 757; | Conservative | 0; | Mismatches 1; | Indels 0; Gaps 0; |
| QY | 1655 | TCCGCGAACATTTTACACGATTTTCAGAGAGCGCTTACTACTGGCTTTCGCGAGACACTGCC | 1714 | | |
| | | | | | |
| DB | 788 | TCCGCGAACATTTTACACGATTTTCAGAGAGCGCTTACTACTGGCTTTCGCGAGACACTGCC | 847 | | |
| | | | | | |
| QY | 1715 | GATGATGCTCGCGTTATGTCTTGTGGGATTACGGATACCCAGATAGCGGGAATGGCAAAAC | 1774 | | |
| | | | | | |
| DB | 848 | GATGATGCTCGCGTTATGTCTTGTGGGATTACGGATACCCAGATAGCGGGAATGGCAAAAC | 907 | | |
| | | | | | |
| QY | 1775 | AGAACGACGCTAGTGGATATAATACGTGGCAACAATAGTCACATAGCGCTGGTGTGGCAAG | 1834 | | |
| | | | | | |
| DB | 908 | AGAACGACGCTAGTGGATATAATACGTGGCAACAATAGTCACATAGCGCTGGTGTGGCAAG | 967 | | |
| | | | | | |
| QY | 1835 | GCAATGTCTTCAACCGAGAGAGAGTCTTACGAAATTTATGACATCTCTTGACGTGGACTAC | 1894 | | |
| | | | | | |
| DB | 968 | GCAATGTCTTCAACCGAGAGAGAGTCTTACGAAATTTATGACATCTCTTGACGTGGACTAC | 1027 | | |
| | | | | | |

```

QY 1895 GTTTTGGTGATCTTTGGCGGTGTGATCGCTATTCTGGCGATGATATCAACAAGTTCTGT 1954
Db 1028 GTTTTGGTGATCTTTGGCGGTGTGATCGCTATTCTGGCGATGATATCAACAAGTTCTGT 1087
QY 1955 TGGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTT 2014
Db 1088 TGGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTT 1147
QY 2015 ACCGACCGCGTGAATTCAGGATAGATGCGAAGGTGCTCCGGCCCTGCTCAACTGCGTT 2074
Db 1148 ACCGACCGCGTGAATTCAGGATAGATGCGAAGGTGCTCCGGCCCTGCTCAACTGCGTT 1207
QY 2075 ATGTACAAATTAAGCTACTACAGATTCCGGGAATTTGAATTTGGACTACAGAGTCCCATCT 2134
Db 1208 ATGTACAAATTAAGCTACTACAGATTCCGGGAATTTGAATTTGGACTACAGAGTCCCATCT 1267
QY 2135 GGATATGATCCGACACGCTAAGCGGTGCTTTGGGAATTAAGGACTTCGATCTGACCTACCTG 2194
Db 1268 GGATATGATCCGACACGCTAAGCGGTGCTTTGGGAATTAAGGACTTCGATCTGACCTACCTG 1327
QY 2195 GAGGAGGCTTACACACAGAACACTTGGCTTCTTCGCATCTATAGGTTGAAGAGCCGCAT 2254
Db 1328 GAGGAGGCTTACACACAGAACACTTGGCTTCTTCGCATCTATAGGTTGAAGAGCCGCAT 1387
QY 2255 GAGTTCAATAGACCATCACTGAAGACCAAGAGAGAACGATTCTCTCCAGCAAACTTCATT 2314
Db 1388 GAGTTCAATAGACCATCACTGAAGACCAAGAGAGAACGATTCTCTCCAGCAAACTTCATT 1447
QY 2315 TCAGAGAAGAACTCTAAGCGTCCGACGAGGCTACATAGCAACCGACCGGTGTTGTTAAG 2374
Db 1448 TCAGAGAAGAACTCTAAGCGTCCGACGAGGCTACATAGCAACCGACCGGTGTTGTTAAG 1507
QY 2375 GGAAGACGAACTTTGAAATTAACCCAAAAA 2412
Db 1508 GGAAGACGAACTTTGAAATTAACCCAAAAAATAA 1545

```

RESULT 4

US-09-270-767-158/c

; Sequence 158, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 158

; LENGTH: 1660

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-158

Query Match 26.6%; Score 643.2; DB 4; Length 1660;

Best Local Similarity 97.3%; Pred. No. 4.8e-170;

Matches 654; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

QY 1742 GATTACGGATACAGATAGCGGAATGGCAACAGAACGACGCTAGTGGATATAATACG 1801
Db 811 GAAATCGTCTAAATGTTCGGGAATGGCAACAGAACGACGCTAGTGGATATAATACG 752
QY 1802 TGGAAACAATAGTCACATAGCGCTGGTGGCAAGGCAATCTTTCAACGAGGAGAGTCC 1861
Db 751 TGGAAACAATAGTCACATAGCGCTGGTGGCAAGGCAATCTTTCAACGAGGAGAGTCC 692
QY 1862 TACGAAATTAATGACATCTCTTACGCTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 1921
Db 691 TACGAAATTAATGACATCTCTTACGCTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 632
QY 1922 GCGTATTCTGGCGATGATATCAACAAGTTCTGTGGATGGTCCGAATTTGCTGAGGAGAG 1981

```

```

Db 631 GCGTATTCTGGCGATGATATCAACAAGTTCTCTGTGGATGGTCCGAATTTGCCGAGGAGAG 572
QY 1982 CATCCCAAGACATTAAGGAAAGCGATTACTTTTACCGACCGCGGTGAATTTCAAGGTAGAT 2041
Db 571 CATCCCAAGACATTAAGGAAAGCGATTACTTTTACCGACCGCGGTGAATTTCAAGGTAGAT 512
QY 2042 GCCGAAGGTGCTCCGGCCCTGCTCAACTGCTTTATGTACAAATTAAGCTACTACGATTC 2101
Db 511 GCCGAAGGTGCTCCGGCCCTGCTCAACTGCTTTATGTACAAATTAAGCTACTACGATTC 452
QY 2102 GGGGAATTAAGTTGAGCTACAGAGTCCATCTGGATATGATCGCACACGTTAACCGCTC 2161
Db 451 GGGGAATTAAGTTGAGCTACAGAGTCCATCTGGATATGATCGCACACGTTAACCGCTC 392
QY 2162 ATTGGGAATTAAGGACTTCGATCTGACCTTACCTTGGAGGAGGCTTACACACAGAACACTGG 2221
Db 391 ATTGGGAATTAAGGACTTCGATCTGACCTTACCTTGGAGGAGGCTTACACACAGAACACTGG 332
QY 2222 CTTGTCGCATCTATAGGTTGAAGAGCCGCTATGTTCAATAGACCATCACTGAAGACC 2281
Db 331 CTTGTCGCATCTATAGTGTGAAGAGCCGCTATGTTCAATAGACCATCACTGAAGACC 272
QY 2282 AAGGAGAGAACGATTCTCCAGCAAACTTCATTTCGAGAAGAACTCTAAGCGTCCGAAG 2341
Db 271 AAGGAGAGAACGATTCTCCAGCAAACTTCATTTCGAGAAGAACTCTAAGCGTCCGAAG 212
QY 2342 GCGTACATAGCAACCGACCGCTTCTTAAAGGAAAAACGAACTTGAATAAACCACAA 2401
Db 211 GCGTACATAGCAACCGACCGCTTCTTAAAGGAAAAACGAACTTGAATAAACCACAA 152
QY 2402 AAAAAA 2413
Db 151 AAAAAAATAA 140

```

RESULT 5

US-09-270-767-15440/c

; Sequence 15440, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15440

; LENGTH: 1660

; TYPE: DNA

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-15440

Query Match 26.6%; Score 643.2; DB 4; Length 1660;

Best Local Similarity 97.3%; Pred. No. 4.8e-170;

Matches 654; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

QY 1742 GATTACGGATACAGATAGCGGAATGGCAACAGAACGACGCTAGTGGATATAATACG 1801
Db 811 GAAATCGTCTAAATGTTCGGGAATGGCAACAGAACGACGCTAGTGGATATAATACG 752
QY 1802 TGGAAACAATAGTCACATAGCGCTGGTGGCAAGGCAATGTCTTCAACGAGGAGAGTCC 1861
Db 751 TGGAAACAATAGTCACATAGCGCTGGTGGCAAGGCAATGTCTTCAACGAGGAGAGTCC 692
QY 1862 TACGAAATTAATGACATCTCTTACGCTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 1921
Db 691 TACGAAATTAATGACATCTCTTACGCTGGACTAGCTTTTGGTATCTTTGGCGGTGTGATC 632
QY 1922 GCGTATTCTGGCGATGATATCAACAAGTTCTGTGGATGGTCCGAATTTGCTGAGGAGAG 1981
Db 631 GCGTATTCTGGCGATGATATCAACAAGTTCTGTGGATGGTCCGAATTTGCCGAGGAGAG 572

```

QY 1982 CATCCAGGACATTAAGGAAAGCGATTACTTACCGACCGCGGTGAATTCAGGCTAGAT 2041
DB 571 CATCCAGGACATTAAGGAAAGCGATTACTTACCGACCGCGGTGAATTCAGGCTAGAT 512
QY 2042 GCCGAAGGTGCTCCGGCCCTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATTC 2101
DB 511 GCCGAAGGTGCTCCGGCCCTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATTC 452
QY 2102 GGGGAATTGAAGTTGGATCAAGAGTCCATCTGGATATGATCGACAGCTAACCCGTC 2161
DB 451 GGGGAATTGAAGTTGGATCAAGAGTCCATCTGGATATGATCGACAGCTAACCCGTC 392
QY 2162 ATTGGGAATAGGACTTGGATCTGACCTACTCGAGAGGCTTACACACAGAACTGG 2221
DB 391 ATTGGGAATAGGACTTGGATCTGACCTACTCGAGAGGCTTACACACAGAACTGG 332
QY 2222 CTTGTTGCATCTATAGGTTGAAGAGCCGATAGTTCAATAGACATCACTGAAGACC 2281
DB 331 CTTGTTGCATCTATAGTGTGAAGAGCCGATAGTTCAATAGACATCACTGAAGACC 272
QY 2282 AAGGAGAGAACGATCTCTCAGCAAACTTCAATTCGAGAAAGAACTTAAGCGTGCAG 2341
DB 271 AAGGAGAGAACGATCTCTCAGCAAACTTCAATTCGAGAAAGAACTTAAGCGTGCAG 212
QY 2342 GGCTACATACAAACCGACCGTGTGTTTAAAGGAAAGAACTTAAGCGTGCAG 2401
DB 211 GGCTACATACAAACCGACCGTGTGTTTAAAGGAAAGAACTTAAGCGTGCAG 152
QY 2402 AAAAAAAAAA 2413
DB 151 AAAAAAAAAA 140

RESULT 6
US-09-270-767-1318/c
; Sequence 1318, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1318
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1318

Query Match 22.9%; Score 552.4; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 7.5e-145;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1107 TCATTGATTCGTTGCGGAGATCAGCCACCACTTGGTCTCTGTTCTTTGATCTGC 1166
DB 560 TCATTGATTCGTTGCGGAGATCAGCCACCACTTGGTCTCTGTTCTTTGATCTGC 501
QY 1167 ACATCCCTGTTGCGGCTTCCAGTGGGAGTGTGTTACTGATCAAGAGATCAACGACG 1226
DB 500 ACATCATGTTGCGGCTTCCAGTGGGAGTGTGTTACTGATCAAGAGATCAACGACG 441
QY 1227 AGCGGTTTTCGTTGCTGATCGCCATAGTCCGTTTACTTCGTTGATGTTGC 1286
DB 440 AGCGGTTTTCGTTGCTGATCGCCATAGTCCGTTTACTTCGTTGATGTTGC 381
QY 1287 GTTTGATTTGACCTCAGCCGCTGTTGATGCTGCGGAGTGGCTTTTCGGAC 1346
DB 380 GTTTGATTTGACCTCAGCCGCTGTTGATGCTGCGGAGTGGCTTTTCGGAC 321
QY 1347 TGTGGATGTTCTCTGCAAGAGATTCTGTTAAGGAATGGGCAGAGCCATTAAGCGAG 1406

DB 320 TGTGATGTTCTCTGCAAGAGATTCTTAACCGAATGGGCACAGCCATAAGCGCAG 261
QY 1407 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAGGCTG 1466
DB 260 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAGGCTG 201
QY 1467 GCAAGCTGAAGCATGTTACTAAGCATGATGCCACAGAGTACTGGGCTAGCTCCCAAC 1526
DB 200 GCAAGCTGAAGCATGTTACTAAGCATGATGCCACAGAGTACTGGGCTAGCTCCCAAC 141
QY 1527 TGAAGAGTATTGTTATTTTGGCCGTTCTAATGCTTTCATGATGTTCCGCTTCCAC 1586
DB 140 TGAAGAGTATTGTTATTTTGGCCGTTCTAATGCTTTCATGATGTTCCGCTTCCAC 81
QY 1587 CGTGGGTGACCAAGCATGCTACTCCAGTCCCTCATTTGCTTTCACACAGCTC 1646
DB 80 CGTGGGTGACCAAGCATGCTACTCCAGTCCCTCATTTGCTTTCACACAGCTC 21
QY 1647 AAGATGATCCCGC 1660
DB 20 AAGATGATCCCGC 7

RESULT 7
US-09-270-767-16600/c
; Sequence 16600, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16600
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-16600

Query Match 22.9%; Score 552.4; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 7.5e-145;
Matches 553; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1107 TCATTGATTCGTTGCGGAGATCAGCCACCACTTGGTCTCTGTTCTTTGATCTGC 1166
DB 560 TCATTGATTCGTTGCGGAGATCAGCCACCACTTGGTCTCTGTTCTTTGATCTGC 501
QY 1167 ACATCCCTGTTGCGGCTTCCAGTGGGAGTGTGTTACTGATCAAGAGATCAACGACG 1226
DB 500 ACATCATGTTGCGGCTTCCAGTGGGAGTGTGTTACTGATCAAGAGATCAACGACG 441
QY 1227 AGCGGTTTTCGTTGCTGATCGCCATAGTCCGTTTACTTCGTTGATGTTGC 1286
DB 440 AGCGGTTTTCGTTGCTGATCGCCATAGTCCGTTTACTTCGTTGATGTTGC 381
QY 1287 GTTTGATTTGACCTCAGCCGCTGTTGATGCTGCGGAGTGGCTTTTCGGAC 1346
DB 380 GTTTGATTTGACCTCAGCCGCTGTTGATGCTGCGGAGTGGCTTTTCGGAC 321
QY 1347 TGTGGATGTTCTCTGCAAGAGATTCTGTTAAGGAATGGGCAGAGCCATTAAGCGAG 1406
DB 320 TGTGATGTTCTCTGCAAGAGATTCTTAACCGAATGGGCACAGCCATAAGCGCAG 261
QY 1407 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAGGCTG 1466
DB 260 CCACCGAAGTGAAGCTGAGGATTCCATTGAGAAAGAACGCTGTACGACAGGCTG 201
QY 1467 GCAAGCTGAAGCATGTTACTAAGCATGATGCCACAGAGTACTGGGCTAGCTCCCAAC 1526
DB 200 GCAAGCTGAAGCATGTTACTAAGCATGATGCCACAGAGTACTGGGCTAGCTCCCAAC 141

QY 1527 TGAAGAGTATTGTTATTTTGGCCGTTCTAATGCTTGAATGATGTTGCTGTCCACTGCA 1586
DB |||||
140 TGAAGAGTATTGTTATTTTGGCCGTTCTAATGCTTGAATGATGTTGCTGTCCACTGCA 81
QY 1587 COTGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTTCCACACAGTC 1646
DB |||||
80 COTGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTTCCACACAGTC 21
QY 1647 AAGATGGATCCCGC 1660
DB |||||
20 AAGATGGATCCCGC 7

RESULT 8

US-09-614-221A-318
; Sequence 318, Application US/09614221A
; Patent No. 6723837
; GENERAL INFORMATION:
; APPLICANT: Karunanandaa, Balasulojini
; APPLICANT: Yu, Jaehyuk
; APPLICANT: Kishore, Ganesh M.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED
; TITLE OF INVENTION: WITH STEROL SYNTHESIS AND METABOLISM
; FILE REFERENCE: 16516.075
; CURRENT APPLICATION NUMBER: US/09/614,221A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/142,981
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 626
; SEQ ID NO 318
; LENGTH: 2157
; TYPE: DNA
; ORGANISM: *Saccharomyces cerevisiae*
US-09-614-221A-318

Query Match 21.7%; Score 523.4; DB 4; Length 2157;
Best Local Similarity 54.8%; Pred. No. 2.2e-136;
Matches 1212; Conservative 0; Mismatches 876; Indels 123; Gaps 4;

QY 105 TGGCTGGCTACAGAGCCTAATCACCTTCGCCATCTGCTAATCGCGTGGCTGGCCGGAT 164
DB 26 TGCTGTGTTTCAGACCATCTCAGCTGCTCATCTTCGTGGCGATTTTGGGGCTGCCA 85
QY 165 TTTCCTCTGCGCTTTCGCGTCATCGTTTCAGTTCGATTATCCATGATTTGATCCGT 224
DB 86 TATCATCACGTTGTTTGCAGTCATCAAAATTCAGTCTATATCCATGATTCGACCCCT 145
QY 225 GGTTCACCTACCGGCCACCGCTACATGTCGAGATGTTGGTACAACTTCTCTCACT 284
DB 146 GGTTCATATAGGGCTACCAATATCTGCTCAACAATTCGTTTACAAATTTTGAAT 205
QY 285 GGTTCGACGCGCGATGTTATCCGCTCGGCAGGATTTGGCGGTACGTTCTATCCG 344
DB 206 GGTTCGACGCGGTACCTGTTACCCCTCGAAGGTTACTGGAGGATTTTATATCTG 265
QY 345 GCCTGATGATAGTTCGCGCGGAATC---CATTTGGCTGTGACGTAACATACCGG 401
DB 266 GTTGTGATGACGATAGTGGCTTATCTGCGCACCCCTCGGCACCTGGTTGGGCTTGCCCA 325
QY 402 TCCATATTCGTGACATCTGGGTGTTCTGCGCGGATCTTCAGTGGCCTGACCTCATCT 461
DB 326 TTGACATCAAGAACGTTTGTGTGCTATTGGCGCACTATTTCTGGGGTACCGCTGGG 385
QY 462 CCACCTACCTGTGACCAAGAGCTGTGTCGCGGCGCCGCTCTTCGCGCGCAGCT 521
DB 386 CGACTTACGAATTTACGAAGAGATTAAGATGCCAGCGCTGGGCTTTGGCTGTGTT 445
QY 522 TCATCGCATCTGCTGCTGCTACATCAGTAGTTCGTTGGCTGATGATGATGATGATGATG 581
DB 446 TTTAGCCATGTCCTGGTTTATATCTAGATCAGTGGCGGGTCTACGATATAGG 505
QY 582 GCATTGCCATATTCGCGCTCGATTCACCTACTTCTGTGGGTGGCTCAGTGAAGACTG 641
DB |||||

DB 506 CCATTGCCATTACATTAAATGGTCACTTTGATGTTTGGATTAAAGGCCCAAAAGACTG 565
QY 642 GATCGGTGTTCTGGTGGCGCGACGCGCTTTGCTACTTCTACATGTTGCTGGCTGGG 701
DB 566 GCTCTATCATGCAAGCAAGCTGTGACGCTTTATTTCTACTTCTACATGTTGCTGGG 625
QY 702 GTGGCTACGTTTCTATCACTCAACCTGATACCCCTGACCTCTTCGCTACTGCTCATATG 761
DB 626 GTGATACGTTTCTATCACTCAACCTGATACCCCTGATACCCCTGATACCCCTGATAC 685
QY 762 GAGGTACTCGCGCGCTCTGCTGAACGCTACAGACCTTCTACATCTCTGGAGTCTGCT 821
DB 686 GCAGATATTCTGTCCTCAAACTGTTCTGCTCTACACCACTTGGTACGCTATTGGAAC 745
QY 822 TCTCCATGACATCCCTTCTGTTGGGATTCACCCGATACGACCCAGTGAACATGGCTG 881
DB 746 CATCCATGACATCCCTTCTGTTGGGATTCACCCGATACGACCCAGTGAACATGGCTG 805
QY 882 CGCTGGAGTGTGTTGCTCTTATGCGCGTGGCCACCTTGGCCATTTGCACTCCGTCG 941
DB 806 CATTGGGTGTTTCTGCTTGTGATTCAGATTGTCGCTTGGTACCTTCGTGAAGGCCAAA 865
QY 942 TGTGCGCAACGAGTTCGGAAGCTGTTTCATGCTGGCGGATTTGCTGGTGGCGTGGCG 1001
DB 866 TCAGCACAGCTAAGTTTAAAGTCATCATGATGTTTCTCTGTTTGTGATCTTGGTCC 925
QY 1002 TCTTTGTGGCGCTGCTGCTGCTCACATGCTGGGGGTTGTGGCCCGCTGGAGTGCAGCT 1061
DB 926 GTGTGTGGGACTTCTGCTTGTGATTCAGATTGTCGCTTGGTACCTTCGTGAAGGCCAAA 985
QY 1062 TCTACTCGCTGTGGGATGCTGCTACGCAAGATCCCAATTCCTCATATGATCCGTTG 1121
DB 986 TTTATTCGTTATGGGATACCAACTACGCAAGATCCCAATTCCTCATATGCTCCGTTT 1045
QY 1122 CGGAGCATCAGCCACCACTTGGTCTCTGTTCTCTTGTGATGCTGACATCTGTTGGTGG 1181
DB 1046 CGGACATCAACCCGTTTGTGGCGCGTCTTCTTGTGATGATGCTGCTGCTGCTGCT 1105
QY 1182 CTTTCCAGTGGAGTGTGATGCTGATCAAGAGATCAACGAGAGCGGTTTTCGTGG 1241
DB 1106 TATTCGCGCGGCTGATTTCTCTACTATTCCTCGACTTGAAGAAGAGCAGCAGTTTGTCA 1165
QY 1242 TGCTGTACGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1301
DB 1166 TCGCTTACTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1225
QY 1302 TCAACCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361
DB 1226 TGACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1285
QY 1362 TGCAAGAGGATTCGCTTGAAGGATGGGACAGCCTAAGCGGACGACCGAAGTGGATG 1421
DB 1286 TGGATTTCAAGACAAAGTGAACGCAATAACGCAATCAAAACCTGCGGCACTACTGG 1339
QY 1422 AAGCTGAGGATTCATTTGAGAAGAAGAGCTGTACGACAAAGCTGGCAAGCTGAAGCATC 1481
DB 1340 ----- 1339
QY 1482 GTACTAGCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1541
DB 1340 -----CCAAATTTGATTTT 1354
QY 1542 TTTTGGCGGTTCTAATGCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601
DB 1355 CCGGATCATTCATCTTTATTTGATCTTTTGTCTTTTGTCTTTTGTCTTTTGTCTTTTGTCTTT 1414
QY 1602 ATGCTACTCAGTCCCTCATTTGCTGCTTTTCCAAACAGTCAAGATGGATCCCGCA 1661
DB 1415 CTGCTACT 1471
QY 1662 ACATTTTAGAGGATTCAGAGGCTTACTACTGCTGCTTTCGAGAACACTGCGGATGATG 1721
DB 1472 CGTTGATCGAGACTTCAGGGAAGCGTACTATTGGTTAAGATGAATCTCTGATGAGGACA 1531

QY 1722 CTCGCGTTATGCTCTGGTGGATTACGGATACAGATAGCGGAATGGCAACACAGCA 1781
DB 1532 GTAGGTTGACGCTGGTGGATTACGGTTACCAATTTGGTGGCATGGCGACAGACCA 1591
QY 1782 CGTAGTGGATAATACTGTGAACAATAAGTACATATGCTGGTGGTGGCAAGCAATGT 1841
DB 1592 CTTTGTGCGATAACAACACGCTGAACAATACTACATCGCCATCGTGTGTAAGCAATGG 1651
QY 1842 CTTCAACCGAGGAGAGCTCTAGCAATATGATCATCTCTTGACGTTGAGTACTGTTTGG 1901
DB 1652 CTTCCCTTGAGAGAAATCTTACCAATTTCTAAAGACATGATGTGATATGCTTGG 1711
QY 1902 TGATCTTTGGCGTGTGATCGGCTATTCTGGCGATGATATCAACAGTTCTGTGGATGG 1961
DB 1712 TCATCTTTGGTGGTCTAATTTGGGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1771
QY 1962 TCCGAATTTGCTAGGAGAGAGATCCCAAGGACATTAAGAAAGCGATTACTTTACGAGC 2021
DB 1772 TCAGATTTAGCGAGGAGATCTGGCAGAGAGATTAAGAGCGTGNATTTCTATACCGAG 1831
QY 2022 CGGTGTAATTCAGGTTAGTCCGAAGGTGCTCCGCCCTGCTCAACTGCTTTATGTACA 2081
DB 1832 AGGAGAAATACAGAGTAGTCAAGGCTTCTGAGACCATGAGAACTCGCTACTTTACA 1891
QY 2082 AATTAAGCTACTACAGATTCGGGGAATTTGAAGTTGACTACAGAGTCCATCTGGATAG 2141
DB 1892 AGATGCTTCAAGATTTCCCAATTTATTCATGTTGGCCAGCACTGACAGAGTGC 1951
QY 2142 ATCCACACAGTAACCGCTCATTTGGGAATAAGACTTCGATCTGACCTACCTGGAGAG 2201
DB 1952 GTCAACAAATGATCACACCATTAG-----ACGTCCACCATTAGACTACTTCGACGAG 2005
QY 2202 CCTACACACAGACACTGCTGTTGCGCATCTATAGGTTGAGAGCGCGATGAGTTCA 2261
DB 2006 TTTTACTTCCGAAATCTGATGTTAGAAATATCAATTAAGAGAGATGATGCCCAAG 2065
QY 2262 ATAGACCATCACTGAGAACCAAGAGAGACGATTCCTCCAGCAACTTCA 2312
DB 2066 GTAGAACTTTGAGGAGAGTGGTGGTGAAGTTAACAGGTCTTCTACGAAACCA 2116

RESULT 9
US-09-270-767-11648
; Sequence 11648, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11648
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-11648

Query Match 14.4%; Score 349.2; DB 4; Length 1386;
Best Local Similarity 61.6%; Pred. No. 1.2e-87;
Matches 558; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

QY 121 CCTAATCACCTTCGCCATCTGCTAATCGCTGGCGGAGATTTCTCTCGCTCTT 180
DB 467 CTTGGTCAAGCTGCGCAATCTCATCTCGGAGCGGTTTATCATTTGACACAGTGT 526
QY 181 GCGCGTATCGTTTCAGTGCATATTCATATGAGTTGATTCGTTCAATACCGGCG 240
DB 527 CTCTGTGCTGGATTCGAAAGCGTAAATCCATGAGTTGATCGGTACTTCAACTACCGCAC 586
QY 241 CACCGCCTACATGTTGTCAGATGTTTGGTACAACTTCTCACTGTTCCGAGCGCGC 300

DB 587 CACGCGTTTCTGGCGGAGCAGGGCTTTTACAAGTTCCACAACCTGGTTTCAATGACCGCG 646
QY 301 ATGGTATCCGCTCGGAGGATTTGGGCGGTACCGTCTATCCGCGCTGATGATTAGTC 360
DB 647 CTGGTATCCCTCGGCGCGCATCATCGGCGGACCACTATATCCCGGCTGATGCTCACTC 706
QY 361 CGCGGAAATCCATTTGGCTGCTCAGCTACTCAACATACCGGTCCATATTTCTGACATCTG 420
DB 707 GCGCGCCCTGTACCCCTGATGTGCTGCTCAATGTGACCATCGACATACGGAACGTTG 766
QY 421 CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACTGCTGACCAA 480
DB 767 CGTCTTCCTGGCGCGCTTCTTCTCTGCTGACCACTGCTGCTGCTGCTGCTGCTGCTG 826
QY 481 GGAGCTGTGCTCGGCGCGCTCTTTCGCGCCACAGCTTTCATCGCACTGCTGCTGCTGCTG 540
DB 827 GGAGATACACAGCACTGAGGCTGGACTGCTGGCGCGCTTTGATATCCATCGTTCCCGG 886
QY 541 CTRACATCAGTAGTCTGGTGGCTGGATCGTACGATACAGAGGCGCATTTGCCATATTTGCGCT 600
DB 887 GTATATCTCTCATCCGTGGCGGATCGTACGAAATGAAGGCTATCGCCATTTTCTGCTGAT 946
QY 601 GCAGTTCACTACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 947 GCTCTTCACTACTTATTTGCTGATCAAGCGGTAAAGACGGGACGATCTTTTGGTGGC 1006
QY 661 CGCAGCGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 1007 TATGTCGGCATTTGGCTACTTCTATATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1066
QY 721 CAACCTGATACCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 1067 TAACCTAATCCCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1126
QY 781 GCTGACCACTACAGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 1127 CTATATGATACAGCAGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1186
QY 841 CGTGGGATTTCAACCGGATACGACCCAGTAAACACATGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 1187 TGTGGGATTTCAACCCATCCAGAGCTCCGAAACACATGCTGCACTGCGAACCCTTTGCGCT 1246
QY 901 CTTATGCGCGTGGCCACCTTGGCGCATTTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 1247 GTGCCAGATTCAGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1306
QY 961 GAAGCTGTTTCATCTGCTGCGCGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 1307 TCTGCTCTTCAAGACGTTGCTTCCAGTGTGTTTCCAGTGTGTTTCCAGTGTGTTTCCAGTGTG 1366
QY 1021 GCTCAC 1026
DB 1367 GCTCAC 1372

RESULT 10
US-09-270-767-12856
; Sequence 12856, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12856
; LENGTH: 914
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-12856

| Best Local Similarity | 73.18; | Pred.No. 2.2e-68; | | | | | | | |
|-----------------------|--------|--|------|------------|------|--------|----|------|-----|
| Matches | 373; | Conservative | 0; | Mismatches | 134; | Indels | 3; | Gaps | 17; |
| QY | 1606 | CPACTCCAGTCCCTCCCAATTCTCTTGCGTTTCCACAACAGTCAAGATGGATCCGCAACAT | 1665 | | | | | | |
| Db | 1 | CTACTCTAGTCCAAAGTGTAGTCTCGGCCTCATACAA--TCATGATGGCACAGGAATAT | 57 | | | | | | |
| QY | 1666 | TTTAGACGATTTACAGAGAGCTTTACTACTGGGTTTCGCAGAACACTGCGGATGATGCTCG | 1725 | | | | | | |
| Db | 58 | CTTAGTGAATTTAGAGAAGCTTACTTTTGGCTAAGGCCAAATACAGATGAACATGCACG | 117 | | | | | | |
| QY | 1726 | CGTTATGCTTGGTGGGATTCAGGATACAGAGTAGCGGGAATGCCAACACAGAACGAGCT | 1785 | | | | | | |
| Db | 118 | AGTAATGTCCTTGGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACTACGTT | 177 | | | | | | |
| QY | 1785 | AGTGATTAATAACGTGGAAACAATAGTCACATAGCGCTGGTGGCAAGGCAATGTCTTC | 1845 | | | | | | |
| Db | 178 | GGTGATTAATAAACCTGGGAATAACAGCCACATAGCACTGGTGGGAAAGCATGTCTTC | 237 | | | | | | |
| QY | 1846 | RACCGAGGAGAGTCCTACCAAAATATGACATCTCTTGACGTGGACTACGTTTGGTGAT | 1905 | | | | | | |
| Db | 238 | TAATGNAACAGCAGCCTATAAAATCATAGGACTCTAGATGAGATATGTTTGGTTAT | 237 | | | | | | |
| QY | 1906 | CTTTGGCGGTGATCGGCTATTCTCGCGATGATATCAACAAGTTCCTGTGGAGTGTCCG | 1965 | | | | | | |
| Db | 298 | TTTTGGAGGGTTATTGGCTATTCTGTGATGATATCAACAAATTTCTCTGGATGGTATG | 357 | | | | | | |
| QY | 1966 | AAATGCTCAGGGAGAGATCCCAAGACACATTAAAGAAAGCGATTACTTTACCGACCCGCG | 2025 | | | | | | |
| Db | 358 | GATAGCTGAAGGAGAACATCCCAAGACATTCCGGGAAGTGACTATTTTACCCACACAGG | 417 | | | | | | |
| QY | 2026 | TGAATTCAGGGTAGATCCGGAAGTGCTCCGCGCTTGCTCACTGCCTTATGTACAAATT | 2085 | | | | | | |
| Db | 418 | AGAAATTCGGTGTACAAAGACAGGATCCCGCTACTTTGTGAAATGCGCTTATGTATAAAT | 477 | | | | | | |
| QY | 2086 | AAGTACTACAGATTCGGGGAATGAGTT | 2115 | | | | | | |
| Db | 478 | GTCACTACTACAGATTTGGGAATGCAGCT | 507 | | | | | | |

```

RESULT 12
US-09-248-796A-3089
  ; Sequence 3089, Application US/09248796A
  ; Patent No. 6747137
  ; GENERAL INFORMATION:
  ; APPLICANT: Keith Weinstock et al
  ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
  ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
  ; FILE REFERENCE: 107196.132
  ; CURRENT APPLICATION NUMBER: US/09/248,796A
  ; CURRENT FILING DATE: 1999-02-12
  ; PRIOR APPLICATION NUMBER: US 60/074,725
  ; PRIOR FILING DATE: 1998-02-13
  ; PRIOR APPLICATION NUMBER: US 60/096,409
  ; PRIOR FILING DATE: 1998-08-13
  ; NUMBER OF SEQ ID NOS: 28208
  ; SEQ ID NO 3089
  ; LENGTH: 867
  ; TYPE: DNA
  ; ORGANISM: Candida albicans
  ; FEATURE:
  ; NAME/KEY: unsure
  ; LOCATION: (25)
  ; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

```

| | | | | |
|-----------------------|--------------|--|----------------|------------|
| Query Match | 10.1% | Score 244.4 | DB 4 | Length 867 |
| Best Local Similarity | 60.1% | Pred. No. 2.2e-58 | | |
| Matches 445 | Conservative | 0 | Mismatches 286 | Indels 9 |
| Caps | 2 | | | |
| QY | 1534 | TATGTGTAATTTGGCCGTTCTAATGTGTTGTATGATGTTGCTGTCCACGTCACGTGGGT | 1593 | |
| Db | 48 | TTGGTTTAACTGACATTTACATTTTACCTTTTCTACTTTGTTTAACTGTTACTCTTGGGT | 107 | |

QY 1594 GACCAGCAATGCTTACTCCAGTCCCTCCATGCTTCTGGCTTCCCAACAGTCAAGTGG 1653
 DB 108 AACATCGATGCTTATTCATCACCACAGTGTGTTAGCATCCAGAAC--CCAGATGG 164
 QY 1654 ATCCGCAACATTTAGACGATTCAGAGAGGTTTACTACTGCTTTTCAGAAACCTGC 1713
 DB 165 CTCACACATATCATTCGATGATATAGAGAAGCTTATCTGTTAAGAAATGAATACACC 224
 QY 1714 CGATGATGCTCGCTTATGCTTCTGGTGGATACAGATACAGATAGCGGGAATGGCAA 1773
 DB 225 AGAAGATGCCAAGTTATGCTTGGCTGGGATATGTTATCAATCGGGGGTATGGCTGA 284
 QY 1774 CAGAACGACGCTAGTGGATTAATATAGTGGAAATAGTCAATAGTCCGCTGGTGGCAA 1833
 DB 285 TAGAACACACATCTTGTATACCAATACATGGAATAACACACATATGGCCACTGTTGGTAA 344
 QY 1834 GGCAATGCTTCAACCGAGGAGAGTCTCAGGAAATATGACATCTTTGACGTGACTA 1893
 DB 345 GGCATGCTTCCCTCGAATGTGCTGATGAATTTTGAGACACACAGATGTTGATTA 404
 QY 1894 CGTTTGGTGAATCTTTGGCGGTGATCGGCTATTCTGGCGATGATATCAACAGTCTCT 1953
 DB 405 TGTCTAGTATATATTTGGAGGGTTATTTGGGTATTCTGCTGATGATATTAACAAATTTCT 464
 QY 1954 GTGATGCTCGAATTCCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGGATTAATT 2013
 DB 465 ATGATGTTAAGATTCCTGAGGTATCTGGCTGATGAATCAAGAAAGAGACTACTT 524
 QY 2014 TACCGACGCGGTGAATTCAGGTAGATGCCAGGTGCTCGCGCTCTCTCACTGCTCT 2073
 DB 525 TACTGACCGAGGAGATATAAGTGGATGAAGATGATCACTGTCGCAATGAAGAATTTCT 584
 QY 2074 TATGTACAAATTAAGTACTACAGATTCGGGGAATGAAGTTGGACTACAGAGGTCCATC 2133
 DB 585 GATGTATAAGTTATCGTATCATAGATTCAGTGAATTTGNGGTAGAGATGGTGA 644
 QY 2134 TGGATATGATCGCACGTAACCGGTCATGCGGATGAGGATTAAGGACTTGCATCGACCT 2193
 DB 645 TAGAGTTAGAAACCAACAAATCCAGCCAAATGAAGTACCGAAATGAATGT-----TGT 698
 QY 2194 GGAGAGGCTTACACACAGACACTGGCTTGTTCATCTATAGGTTGAAGAAGCGCA 2253
 DB 699 TGAAGAGCCTTACATCAGAAATTTGGATTTGAGAAATTACAAATTAAGATTTGA 758
 QY 2254 TGAGTTCAATAGACCATCAC 2273
 DB 759 TAATGTTGGTAGATTTAC 778

RESULT 13

US-09-328-111-332
 ; Sequence 332, Application US/09328111
 ; Patent No. 6262333
 ; GENERAL INFORMATION:
 ; APPLICANT: Endege, Wilson O.
 ; APPLICANT: Steinmann, Kathleen E.
 ; APPLICANT: Astle, Jon H.
 ; APPLICANT: Burgess, Christopher C.
 ; APPLICANT: Bushnell, Steven E.
 ; APPLICANT: Carroll III, Eddie
 ; APPLICANT: Catino, Theodore J.
 ; APPLICANT: Derti, Adnan
 ; APPLICANT: Ford, Donna M.
 ; APPLICANT: Lewis, Marcia E.
 ; APPLICANT: Monahan, John E.
 ; APPLICANT: Schlegel, Robert
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 ; FILE REFERENCE: CCD-257 (US)
 ; CURRENT APPLICATION NUMBER: US/09/328,111
 ; CURRENT FILING DATE: 1999-06-08
 ; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10
 ; NUMBER OF SEQ ID NOS: 850
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 332
 ; LENGTH: 616
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(616)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-09-328-111-332

Query Match 7.2%; Score 173; DB 3; Length 616;
 Best Local Similarity 64.0%; Pred. No. 1.9e-38;
 Matches 260; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 533 GTGCTGGCTACATCAGTAGTGGTGGCTGGATCGTACGATAACGAGGCGATTGCCATA 592
 DB 5 GTACACAGGCTACATATCTCGGTCACTAGTCTGGATCCCTTGATATATGAAGCATTTGCTATT 64
 QY 593 TTGCGCCCTCAGTTTCACTACTTCTGTTGGTGGCTCAGTGAAGACTGATCCGTTTC 652
 DB 65 TTGCACTTCACTTACATATTTATGGTAAATCTGTAAAACTGGGTGAGTTT 124
 QY 653 TGTGCGGCGCAGCGCTTTGCTTCTTCTACATGTTGCTGGCTGGGTGGCTAGCTG 712
 DB 125 TGGACAATGTGCTGCTTATCTTATTTCTATATGTTCTCTGCTGGGTGGTTATGTA 184
 QY 713 TTATCATCAACTGATACCTCTGCACTCTTCTGTTCTGTTCTGTTCTGTTCTGTTCTG 772
 DB 185 TTATCATCAACTTATTTCCACTGCACTTATTTGTTGTTGTTGTTGTTGTTGTTGTTG 244
 QY 773 CGCGTCTCTGACAGGCTACAGACCTTCTACATCTCTGGGACTGCTGTTCTCCATGCG 832
 DB 245 AABAGAGTCTACATAGCATATAGCATTTCTACATGTTGGGTTTAAATATATCAATGCG 304
 QY 933 ATCCCTTCTGTTGGATTCACCGGATAGCAGCAGTGAACACATGCTGGCTGGAGTG 892
 DB 305 ATACCTTTTGTGGGATTCAGCAATCAAGAACAGTGAACACATGGCAGCTTGCAGTGC 364
 QY 893 TTGTGCTCTTATGGCGTGGCCACCTTGGCCATTGTCAGTCCG 938
 DB 365 TTGCACTCTGCAAGCTTAATCTTTCTTTCAGTATCTGAGAACCG 410

RESULT 14

US-09-270-767-27262
 ; Sequence 27262, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7328-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 27262
 ; LENGTH: 503
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 ; US-09-270-767-27262

Query Match 7.1%; Score 171.2; DB 4; Length 503;
 Best Local Similarity 59.4%; Pred. No. 5.5e-38;
 Matches 290; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 539 GGCTACATCAGTAGTGGTGGCTGGATCGTACGATACGAGGCGATTGCCATATTCGCC 598
 DB 2 GGGTATATCTTCGATCCGTTGGGATCGTACGATCAATGAAGCATCGCATTTTTCGC 61
 QY 599 CTGCACTTCACTTCTCTGTTGGTGGCTGCTGAGTGAAGACTGATCCGTTCTGTTCTG 658

```
Db 62 ATGCTCTTCACTACTATTGTTGGATCAAGCGGTAAAGACGGGACGATCTTTGGTCG 121
QY 659 GCGGACGCGCTTTGCTCTACTTCTACATGTTGTCGGCTGGGCTGCTAGCTTTCATC 718
Db 122 GGTATGTGGGCAATGGGCTACTTCTATATGTTCTCTCTGGTGGGCTATGCTTTCCTG 181
QY 719 ATCAACCTCATACCCCTGCAAGCTCTTCGTACTGTCTCATATTGGGCAAGTACTCGCCCGT 778
Db 182 ATTAACCTAATCCGCTGCAAGTGTGGGCTGATGATCAACCGAAGCTTCTCGCACAGG 241
QY 779 CTGCTGACAGCTACAGACCTTCTACATCTGGAATGCTGTTCTCGATGCAATCCCC 838
Db 242 ATCTACATAGCATACAGACGCTTACTGCTGCGGCAACCATTCCTGCGATGCAATCTCG 301
QY 839 TTGCTGGGATTCACACCCATAGCACCAAGTGAACACATGGCTGGGCTGAGTGTGTG 898
Db 302 TTGTGGGATTCACACCCATCCAGAGCTCCGAACACATGCTGGCAC-TGGAAACCTTTGGC 361
QY 899 CTCCTTATGGCGTGGCCACCTTGGCCATTTGCAATCCGCTGCTGTGCGCAACGATTC 958
Db 362 CTGTGCCAGATTCAAGCTTTTGTGACTATCTGGCTCGGCAATTCACAAAGGATCACTTC 421
QY 959 CGGAAGCTTTTCATCGTGGCGGATTTGCTGTGGCGCTTGGCTGTTGTGGCGCTCGTG 1018
Db 422 GATCTGCTCTCAAGACGTTGTTCCAGTGTTTTGACTGTGGTGTTCGTCGTGGGTACC 481
QY 1019 GTGCTCAC 1026
Db 482 CTGCTCAC 489
```

RESULT 15

```
US-09-385-982-213
; Sequence 213, Application US/09385982
; Patent No. 6282334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-213
```

```
Query Match 6.2%; Score 150.2; DB 3; Length 487;
Best Local Similarity 57.3%; Pred. No. 4.2e-32;
Matches 272; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 797 ACCTTCTACATCTCGGACTGCTGTCTCCATGCAAGATCCCTTCGTGGGATTCACACCG 856
Db 1 ACTGTTTACTGCTGGGCACTACTTCTATGCAAGATCTCTTTGTGGGTTTCCAGCCT 60
QY 857 ATACGCACAGTGAAACATGCTGGCTGGGAGTGTGTCCTTATGGCGGTGGCC 916
Db 61 GTCCTTTATCAGAGCACATGGCAGCCTTTGGGTCCTTTGGTCTCTGCCAGATCCATGCC 120
QY 917 ACCTTGGCCATTTGCGATCCGCTGTGCGCAACGAGTTCCGGAAGCTGTTTCATGCTC 976
Db 121 TTGTGGATTACCTGCGCAGCAAGTTGAATCCACAAATTTGAAGTCTTTTCCGGAGC 180
```

```
QY 977 GCGGATTTGCTGGTGGGCTTGGCGTCTTTGTGGCCGCTGCTGTGTCACCATGCTGGGC 1036
Db 181 GTCATCTCTCTGGTAGGCTTTGTCTCTTCTCACCGTGGAGCTCTCTCATGCTGACAGGA 240
QY 1037 GTTGGGCCCCGTGGAGTGAAGCGTTCTACTGCTGTGGGATACTGGCTACGCCAAGATC 1096
Db 241 AAAATATATCTCCCTGACGGGCGTTTCTACTCTCACTGCTGGATCCCTCTTATGCTAAGAAC 300
QY 1097 CACATTCCCATCATTTGATCGTGTGGAGCATCAGCCCAACCATTTGGTTCTCGTTCTTC 1156
Db 301 AACATCCCATCATTTGCTTCTGTGCTGTGAGCATCAGCCCAACCATTTGCTCTCATACTAT 360
QY 1157 TTTGATCTGACATCTGCTGTGGCGCTTCCAGTGGGAGTGTGCTACTGCATCAGCAG 1216
Db 361 TTTGACCTGCACTCTCTGCTTTCATGTTTCCAGTTGGCCTCTATTACTGCTTTAGCAAC 420
QY 1217 ATCAACGACGAGCGGTTTTCGTGGTGTGCTGTACGCCATCAGTGGGTTTACTTCG 1271
Db 421 CTGTCTGATGCCCGGATTTTATCATGATGTTGTGACCAAGCATGTACCTCG 475
```

Search completed: December 15, 2004, 11:16:18
Job time : 196.945 secs

Blank
page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 07:01:53 ; Search time 1208.01 seconds
(without alignments)
11031.546 Million cell updates/sec

Title: US-10-028-384-7

Perfect score: 2417

Sequence: 1 tctagcgaagaagtgtcg.....ccaaaaaataaaaaaa 2417

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 2417 | 100.0 | 2417 | 15 | US-10-028-384-7 |
| 2 | 960.6 | 39.7 | 2710 | 15 | Sequence 7, Appli |
| 3 | 955.8 | 39.5 | 2481 | 15 | Sequence 3, Appli |
| 4 | 636 | 26.3 | 3046 | 18 | Sequence 1, Appli |
| 5 | 635.4 | 26.3 | 2332 | 15 | Sequence 95, Appli |
| 6 | 632.8 | 26.2 | 2472 | 15 | Sequence 7139, Ap |
| 7 | 632.8 | 26.2 | 2472 | 15 | Sequence 112, App |
| 8 | 632.8 | 26.2 | 2472 | 15 | Sequence 11, Appl |
| 9 | 632.8 | 26.2 | 2472 | 15 | Sequence 742, App |
| 10 | 631.2 | 26.1 | 3093 | 18 | Sequence 742, App |
| 11 | 631.2 | 26.1 | 3094 | 15 | Sequence 92, Appl |
| 12 | 628.6 | 26.0 | 5404 | 18 | Sequence 9, Appli |
| | | | | | Sequence 99, Appli |

| | | | | | | |
|----|-------|------|------|----|----------------------|--------------------|
| 13 | 628.6 | 25.0 | 5927 | 18 | US-10-417-375-97 | Sequence 97, Appl |
| 14 | 620.2 | 25.7 | 2466 | 16 | US-10-320-797-2305 | Sequence 2305, Ap |
| 15 | 619 | 25.6 | 2839 | 18 | US-10-425-115-150745 | Sequence 150745, A |
| 16 | 615 | 25.4 | 2681 | 16 | US-10-425-114-14408 | Sequence 14408, A |
| 17 | 612.4 | 25.3 | 2779 | 17 | US-10-437-963-99904 | Sequence 99904, A |
| 18 | 562 | 23.3 | 1848 | 15 | US-10-128-714-2139 | Sequence 2139, Ap |
| 19 | 523.4 | 21.7 | 2157 | 9 | US-09-801-368-387 | Sequence 387, App |
| 20 | 523.4 | 21.7 | 2157 | 18 | US-10-793-639-318 | Sequence 318, App |
| 21 | 521.8 | 21.6 | 2733 | 15 | US-10-028-384-5 | Sequence 5, Appli |
| 22 | 519.8 | 21.5 | 2256 | 15 | US-10-032-585-6323 | Sequence 6323, Ap |
| 23 | 497 | 20.6 | 1969 | 15 | US-10-128-714-1139 | Sequence 1139, Ap |
| 24 | 497 | 20.6 | 2603 | 15 | US-10-128-714-6139 | Sequence 6139, Ap |
| 25 | 497 | 20.6 | 3969 | 15 | US-10-128-714-139 | Sequence 139, App |
| 26 | 497 | 20.6 | 4603 | 15 | US-10-128-714-5139 | Sequence 5139, Ap |
| 27 | 467 | 19.3 | 1209 | 15 | US-10-106-698-330 | Sequence 330, App |
| 28 | 467 | 19.3 | 1209 | 16 | US-10-264-237-412 | Sequence 412, App |
| 29 | 466.4 | 19.3 | 1543 | 10 | US-09-574-879-133 | Sequence 133, App |
| 30 | 466.4 | 19.3 | 1543 | 10 | US-09-305-736-133 | Sequence 133, App |
| 31 | 466.4 | 19.3 | 1543 | 10 | US-09-818-683-133 | Sequence 133, App |
| 32 | 466.4 | 19.3 | 1543 | 11 | US-09-818-683-133 | Sequence 133, App |
| 33 | 466.4 | 19.3 | 1543 | 16 | US-10-621-401-133 | Sequence 133, App |
| 34 | 430.2 | 17.8 | 1828 | 10 | US-09-945-527-62 | Sequence 62, Appl |
| 35 | 403 | 16.7 | 2660 | 16 | US-10-264-049-630 | Sequence 630, App |
| 36 | 382.6 | 15.8 | 3141 | 18 | US-10-425-115-130787 | Sequence 130787, A |
| 37 | 373.8 | 15.5 | 2882 | 16 | US-10-320-797-1305 | Sequence 1305, Ap |
| 38 | 373.8 | 15.5 | 4738 | 16 | US-10-320-797-305 | Sequence 305, App |
| 39 | 357.4 | 14.8 | 3197 | 16 | US-10-424-599-111541 | Sequence 111541, A |
| 40 | 311.6 | 12.9 | 1728 | 16 | US-10-424-599-122476 | Sequence 122476, A |
| 41 | 310.8 | 12.9 | 1114 | 16 | US-10-296-115-629 | Sequence 629, App |
| 42 | 291.2 | 12.0 | 2244 | 17 | US-10-437-963-48342 | Sequence 48342, A |
| 43 | 279.8 | 11.6 | 887 | 16 | US-10-424-599-77697 | Sequence 77697, A |
| 44 | 247.4 | 10.2 | 430 | 9 | US-09-878-178-717 | Sequence 717, App |
| 45 | 247.4 | 10.2 | 430 | 13 | US-10-046-935-717 | Sequence 717, App |

ALIGNMENTS

RESULT 1
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication NO. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27
; RELEVANT RESIDUES: (1)..(2417)
US-10-028-384-7

| | | | | |
|-----------------------|--------------|--|------------|--------------|
| Query Match | 100.0% | Score 2417; | DB 15; | Length 2417; |
| Best Local Similarity | 100.0% | Pred. No. 0; | | |
| Matches 2417; | Conservative | 0; | Mismatches | 0; |
| | | | Indels | 0; |
| | | | Gaps | 0; |
| QY | 1 | TCTAGCGAAGAAGTGTGCTTGCATTTCAGATCGGTATAATTTTCAGTTACTGCTG | 60 | |
| Db | 1 | TCTAGCGAAGAAGTGTGCTTGCATTTCAGATCGGTATAATTTTCAGTTACTGCTG | 60 | |
| QY | 61 | GAATTTGGGACATGATCGGACCGGAGATGCTCTGAACAGCAAGTGGCTGCTACAGCAG | 120 | |
| Db | 61 | GAATTTGGGACATGATCGGACCGGAGATGCTCTGAACAGCAAGTGGCTGCTACAGCAG | 120 | |

Db 2281 CAAAGGAGAAAGGATTCCTCCAGCAAACTCATTTTCGAGAAAGAACTCTAAGCGTCGCA 2340
Qy 2341 GGCTACATACGAAACCGACCGGTTGTTGTTAAAGGAAAGAAAGAACTGTAATTAACCCCA 2400
Db 2341 GGCTACATACGAAACCGACCGGTTGTTGTTAAAGGAAAGAAAGAACTGTAATTAACCCCA 2400
Qy 2401 AAAAAAAAAAAAAAAAAA 2417
Db 2401 AAAAAAAAAAAAAAAAAA 2417

RESULT 2

US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Kevin
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)...(2469)
US-10-028-384-3

Query Match 39.7%; Score 960.6; DB 15; Length 2710;
Best Local Similarity 64.8%; Pred. No. 7, 4e-285;
Matches 1495; Conservative 0; Mismatches 794; Indels 18; Gaps 4;
Qy 102 AGTGCTGGCTACAGCAGCCTAATCAGCTGGCCATCTGCTAATGCGCTGGCTGGCGG 161
Db 258 AGCGGCGGGTGGAGTGGTGGCTCTCTCCATCATCTCTCTCTGGCTGGCTGGCGG 317
Qy 162 GATTTTCCTCTCGCTCTTCGCGCTCATCCGCTTCGAGTGCATATCATGAGTTGATC 221
Db 318 GCTTCAGCTCGCGCTCTTCGCGCTCATCCGCTTCGAGGATCATCCAGGTTCCACC 377
Qy 222 CGTGGTTCACTACCGGCGCACGGCTACATGGTGCAGAAATGGTTGGTACAACTCTCA 281
Db 378 CGTGGTTCACTACATAGATCAACATCATCTTCGATCTCATGGATTTCTAGATTTCTAA 437
Qy 282 ACTGGTTCCAGCAGCGCGCATGGTATCCGCTCGGCGGATTTGGGGGGTACCGCTCATC 341
Db 438 ATTGGTTTATGAAGAGCATGTATCCACCTCGGAGAAATAGTGGTGGCACCGTTTACC 497
Qy 342 CCGGCTGATGATACGTCGGCGGAATCCATTGCTGCTGCAGTACTCAACTACCGG 401
Db 498 CAGGGTTGATGATAACAGCTGGCCCTTATTCATTGGATTTAAATACATTTGAACATAACAG 557
Qy 402 TCATATTCGTCGACATCTCGGTTTCCTGGCGCCGATCTTCAGTGGGCTGACCTCCATCT 461
Db 558 TTCACATAAGAGATGTGTGTTATTCCTTTGCACCAACTTTTAGCGGCTTTACATCCATAT 617
Qy 462 CACCTTACTCTGTACCAAGAGCTGTGGTCCGCGGGCGCGGCTCTTTCGCGCCAGCT 521
Db 618 CTACGTTCTGTCTAATAGAGAACTGTGGAACCAAGAGAGAGAACTTCTAGCTCTTGGCT 677
Qy 522 TCATCGCCATCGTGGCTGCTACATCAGTAGGTGGTGGCTGGATCGTACGATAACGAGG 581
Db 678 TCATCTGCTATCGTACCGGGTACATATCTCGGTGAGTGGCGGATCTTGTATGATGAAG 737
Qy 582 GATTCGCCATATTCGCCCTGAGTTCCACTACTCTCTCTGCTGGGTGGCTCAGTGAAGACTG 641

Db 738 GCATTTGCCATTTTTCGCGCTTCAGTTCACTTACTACTTATCGGTAAAGTCTGTGAAGACCG 797
Qy 642 GATCCGCTGTCTGGTGGCGCCAGCGCTTGTCTTACTTCTTCTACATGGTCTCGCGCTGGG 701
Db 798 GGTCTGTCTCTGGACAAATGTGCTGTGTCATATTTCTACATGCTCTCTCGCTGGG 857
Qy 702 GTGGCTACGCTGTTCATCATCACTGATACCCCTGACGCTTCTGCTACTGCTCTATTATGG 761
Db 858 GAGGTATGTGTTCATCATCACTCATCCCTCTCCATGTGTGTGTGCTGCTGCTGATGC 917
Qy 762 GCAGGTACTCGCGCGCTGCTGCTGACCACTACAGCACTTCTACATCTCTGGGACTGCTGT 821
Db 918 AGAGGTACAGCAAGAGAGTCTACATAGCATATAGCACTTTGTACATTTGCGGTTAATAT 977
Qy 822 TCTCCATGAGATCCCTTCTGGGATTCNACGATACGACCACTGACAGTGAACATGCGCTG 881
Db 978 TATCCATGAGATACCTTTTGTGGATTTTCAGCAATCAGAAAGAGGAGCAATGGCAG 1037
Qy 882 CGCTGGGAGTGTGTGCTCTCTTATGCGCGTGGCCACTTTCGCGCAATTTGCACTCGGTCG 941
Db 1038 CTGCAAGGTCTTTGCGCTGCTGCAAGCTTACGCTTTTTCGAGTATCTGAGAGACCGGT 1097
Qy 942 TGTCCGCAACGAGTTCCGGAAGCTGTTCATCTGCGCGGATTCGCTGGTGGGCTGGCG 1001
Db 1098 TGACAAAACAGGAGTTCAGACCTTTTCTTTTGGGTGCTCTCACTAGTGCAGGCGCTG 1157
Qy 1002 TCTTTTGTGGCGCTGCTGCTGCTCACTGCTGGCGGTTGTGGCGGCTTGGAGTGGAGCT 1061
Db 1158 TGTTCCTTAGTGTCTATCTGACATACACAGTTTATTTGACCACTGAGTGGCAGGT 1217
Qy 1062 TCTACTCTGCTGGGATCTGCTACCGCAAGTTCGATCTGCGCGGATTCGCTGGTGGGCTGG 1121
Db 1218 TTTTACTACTATGGGATCTGGGTATGCAAAATACACATTCCAATTAATTTGCACTAGTGT 1277
Qy 1122 CGAGCATCAGCCACCACTTGGTTCTCGTTCTCTTTGATCTGCACTCTGCTGGTGGC 1181
Db 1278 CTGAACATCAGCTACGACATGGGTGCTTCTTCTTTGATCTACATATTTCTGTATGTA 1337
Qy 1182 CTTTCCGAGTGGGATGTGTGTAATGCAATCAAGAGATCAACGAGCGGCTTTCGTTGGTGG 1241
Db 1338 CTTTCCGAGCGGCTATGGTTCTGCAATCAAAATATCAACGATGAAGAGTATTTGTGCG 1397
Qy 1242 TGCTGTACGCCATCAGTGGCGGTTTACTTCTGCTGTGTGATGGTGGCTTTGATCTGACCC 1301
Db 1398 CTCTGTATGCGATCAGTGTCTGTGTACTTTGCGGAGTGTGGTGGGCTGATGCTGACTC 1457
Qy 1302 TCAGCGCGGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361
Db 1458 TGACCGCGGTGCTGTGATGCTGCTGCGCCATTCGCGCTTCTCCAAATGTTTTCGACACTAT 1517
Qy 1362 TGCAAGAGGATTCGCTTAAAGCAATGGGCAAGCCATTAAGCGAGCCACCGAAGTGGATG 1421
Db 1518 TGGGGATGA-----CATGAAAAGGAAAAAACCACCTCTGGAG-----GACAGCAGTG 1565
Qy 1422 AGCTGAGGATTCATTTGAGAAAGAGAGCTGTACGACAGGCTGGCAAGCTGCAAGCATC 1481
Db 1566 ATGAGGATGACAAAGAAACCCAGGAACTTTGTATGACAAAGGAGGTAAGTGAAGGAGC 1625
Qy 1482 GTACTTAAGCATGATGCGCCAGCAGGATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1541
Db 1626 ATGTGACAGCAAGAGAAACCTGAGAGGCTTTGGGCCCCCAACATCAAAAGCATTTGTA 1685
Qy 1542 TTTTGGCGGCTTAACTGCTGTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601
Db 1686 CCACTGCT 1745
Qy 1602 ATGCTACTTCCAGTCCCTCCATTTGCTTGGCTTTTCCAAACAGTGAAGTGGATCCCGCA 1661
Db 1746 ACGCTACTCTCAGTCCAAAGTGTGTCTCTGCTCTTACAA---TCATGATGGTACAGGA 1802
Qy 1662 ACATTTTGAAGATTCAGAGAGCTTACTACTGGCTTTTCGAGAAACACTGCGGATGATG 1721

Db 1803 ATATATTAGATGATTTAGAGACCGTACTTTTGGCTGAGACAAACACGAGTACACG 1862
 Qy 1722 CTCGGTTATCTCTGGTGGGANTTACCGATACAGATAGCGGAATGGCAAAACAGACGA 1781
 Db 1863 CCCGGGTATCTCGTGGTGGGACTACGGCTATCAGATTTGCTGCATGGCCAAACAGACCA 1922
 Qy 1782 CGCTAGTGGATAAATACGTGGAAACATAGTACATAGCGCTGGTGGTGGCAAGCAATGT 1841
 Db 1923 CTCGTGCTGATACAAACCTGGAAACACAGCACATCGACTGGTGGCAAAAGCTATGT 1982
 Qy 1842 CTTCAACCGAGGAGAGTCTTACGAATATGACATCTCTTGGAGTGGACTAGTTTGG 1901
 Db 1983 CTTCAATGAAACCGCCCTTATAAATCATGAGTCCCTTGTGATGTCGATTAATGTGTGG 2042
 Qy 1902 TGATCTTTGGGGTGTGATCGCTATTCTCGCGATGATATCAACAGTTTCTTGTGGATGG 1961
 Db 2043 TTAATTTCCGAGGAGTGTATGGCTATTCCGGGACGATATCAACAGTTTCTTGGATGG 2102
 Qy 1962 TCCGAATTCGTAGGAGAGCATCCCAAGGACATTAAGGAAGCGATTACTTTACCGACC 2021
 Db 2103 TCAGGATAGCTGAAGGGGAGCATCCCAAGGACATCCCGGAAGGTGACTATTTTCACCCAGC 2162
 Qy 2022 GCGGTGAATTCAGGCTAGATCCCAAGGTGCTCCGGCCCTGCTCAACTGCTTATGTACA 2081
 Db 2163 AGGAGAGTTCGAGTAGACAAAGCTGGTCTCTTACTCTGTAACTGCTTATGTATA 2222
 Qy 2082 AATTAAGCTACTACAGATTCGGGAATGAAGTTGACTACAGAGTCCATCTGGATATG 2141
 Db 2223 AATGTCTATCTACAGATTTGGAGAAATGAGCTAGATTTTCGCACTCCCGAGGCTTGG 2282
 Qy 2142 ATCGCACACGTAACCGCTCATTCGGGAATGAAGCTTCGATCTGACCTACCTGGAGAGG 2201
 Db 2283 ACCGAACAGTAACTGCTGAGATTGGAATTAAGACATTAATCAAGCATTTGGAGGAG 2342
 Qy 2202 CTTACACACAGACACTGGCTTGTTCGCTATCTATAGGCTGAAGAGCCGATAGTTCA 2261
 Db 2343 CTTTATACATCAGACACTGGCTTGTTCAGGATATATAAGTGAAGACACCTGACACAGGG 2402
 Qy 2262 ATAGACACTACTGAAGACCAAGAGAGACGATTCCTCCAGCAAACTTCATTTCGAGAA 2321
 Db 2403 AGACACTAGTTCACAAACCTCGATCCACACATCGTCCCAACAGAGTATTTGTCAA 2462
 Qy 2322 AGAAGCTTAAGCTGCAAG---GGCTACATACGAACCGACCGGTTGTGTATAGGAA 2378
 Db 2463 AGAAGACTACTAAAGGAGCGTGGCTACGTTAAAAATAGCTAGTGTGTTTAAAGAGGCA 2522
 Qy 2379 AACGAACCTTGAAATAAACCCAAAAA 2405
 Db 2523 AGAAGACCTTAAGAGACTGTTTAA 2549

RESULT 3
 US-10-028-384-1
 ; Sequence 1, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREULT, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 2481
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2481)
 ; OTHER INFORMATION:

US-10-028-384-1
 Query Match 39.5%; Score 955.8; DB 15; Length 2481;
 Best Local Similarity 64.8%; Pred. No. 2.1e-283;
 Matches 1489; Conservative 0; Mismatches 792; Indels 18; Gaps 4;
 Qy 102 AGTGCTGGCTACAGACGCTTAACAGCTTGGCATCTCTTAATCGCTGGCTGGCCG 161
 Db 191 AGCGGCTGGGTGGGACGCTTCTCTCTTCCACATCTCTTCTGGCTGGCTGGCCG 250
 Qy 162 GATTTTCCTCTCGCTCTTTCGCGCTATCCGTTTCGAGTCGATTAATCAATGATTTGATC 221
 Db 251 GCTTCAGCTCGGCGCTCTTTCGCGCTATCCGCTTCAAGACATCATCCAGAGTTGAC 310
 Qy 222 CGTGGTTCAATACCGGCGCCACCGCTACATCTGCTGCAAGATGGTTGGTCAACATCTTCTCA 281
 Db 311 CGTGGTTTAACTATAGATCAACACATCATCTTGCATCTCATGGTTCATGAATTTTAA 370
 Qy 282 ACTGGTTTCAGAGCGCGCATGGTATCCGCTCGGAGGATTTGGCGGTTACCGTCTATC 341
 Db 371 ATTGGTTTATGAAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGCTACTGTTTACC 430
 Qy 342 CCGGCTGATGATTAACGCTCGGCGCAATCCATTTGGCTGCTGCAGCTACTCAACATACCGG 401
 Db 431 CAGGTTGATGATTAACCGCTGCTTATTCATTTGAATTTTAAATACATTGAACATACTG 490
 Qy 402 TCATATTCGTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461
 Db 491 TACATATAAGAGACCTGATGTTGTTTCTTTCGCAACATTTTAGCGGCTTACATCTATAT 550
 Qy 462 CCACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
 Db 551 CTACTTCTGCTTACAGAGACCTTTGAAACCAAGGAGCAGGACTTTTAGCTGCTTGT 610
 Qy 522 TCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
 Db 611 TTATTGCTATTGACAGGCTACATATCTCGGTCAGTAGCTGGATCTTTTGAATAATGAG 670
 Qy 582 GATTCCTCATATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
 Db 671 GCATTGCTATTTTGCACCTTACATATCTTATTTATGGTAAATCTGTAATAACTG 730
 Qy 642 GATCCGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701
 Db 731 GGTCACTTTTGGCAATATGCTGCTGCTTATCTATTTCTATATGGTCTGCTGCTGCTGCT 790
 Qy 702 GTGGCTACGTTTCTATCATCACTGATACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
 Db 791 GTGGTTATGTTTATCATCACTTATTTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 850
 Qy 762 GAGGTACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 821
 Db 851 AGAGATACAGCAAAAGAGCTTACATAGCATATAGCATTCTTACATTGTTGGTTTAAATAT 910
 Qy 822 TCTCCATGAGATCCCTTCTGTTGGATTTCCAAACGATACGACCAAGTGAACACATGCTG 881
 Db 911 TATCAATGAGATACCTTTTGGGATTTCCAGCAATCAGAAACAGTGAACACATGSCAG 970
 Qy 882 CGCTGGAGTGTGTTGCTCTTATGGCGTGGCAGCTTGGCCATTTGCACTGCTGCTGCTGCT 941
 Db 971 CTGAGGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1030
 Qy 942 TGTTCGCGCAACGAGTTTCGGAAGCTGTTTCATGCTGCGGAGTTCCTGCTGCTGCTGCTGCT 1001
 Db 1031 TAAACAAAACAGAGTTCCAGACCTTTCTTTTGGGTGTTATCACTAGCTGACGCTGCTGCT 1090
 Qy 1002 TCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1061
 Db 1091 TGTTCCTTAGTGTCTATTTGACTTATACAGGTTACATTGCAACATGAGTGGAGTGGAGT 1150
 Qy 1062 TCTACTCGCTGTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
 Db 1151 TTTATTCTTTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1210

1122 CGGAGCATCAGCCACCACTGGTTCTCGTCTCTTTGATCTGCAATCCTCGTGTGG 1181
1211 CTGAGCATCAACCTACGACTGGGTGCTTTCTTTCTTTGATCTACATATCTTTGATGTA 1270
1182 CTTCCAGTGGAGTGTGGTACTGATCAAGAGAGATCAACGAGCGGCTTTCTGTGG 1241
1271 CTTCCAGCGAGCGCTTTGGTTCTGCTATCAAAATATCAAGATGAAGAGTATTTGTTG 1330
1242 TGCTGTAGCCATCAGTGGGTTTACTTCGTGTGTGTGATGGTGGCTTTGATGTGACCC 1301
1331 CTCTATATGCAATCAGTCTCTCTCTTTGCTGAGTGTGTTGGTCTGATGTGACTT 1390
1302 TCACCGCGGTGTGTGATGCTGTGCGCGAGTGGCTTTTGGGACTGTTGGATGTGTTC 1361
1391 TGACTCCAGTGTGTGTATGCTGTGCAATTTGCTTTTCAAAATGTTTTTGAGCACTATT 1450
1362 TGCAAGAGGATTCGTCTAAGCGAATGGGCACAGCCATAAGCGCAGCCACCGAAGTGGATG 1421
1451 TGGGGATGA-----CATGAAAGGGAATCCACTGTGGAG-----GACAGCAGTG 1498
1422 AAGCTGAGGATTCATGTGAGAGAGAGCTGTGTAGCAAGAGCTGGCAAGCTGAAGATC 1481
1499 ATGAGGATGACAAAAGAAACCAAGAAATTTGTATGATAAGCGCAGGTAAAGTGAGGAAAC 1558
1482 GTACTAAGCATGATCCCGCAGCAGGATCTGCGCTCAGCTCAACCTGCAAGATTTGTTA 1541
1559 ATGCACTGNACAGGAAAACACTGNAGAGGATTTAGGCCCTAATATATAAAGATTTGTCA 1618
1542 TTTTGGCGTTCTAANGCTGTGTGATGTTGCTGCTCCACTGCACTGCGGTGACGAGCA 1601
1619 CCATGTTGATGCTGATGTTGATGATGTTTGTGCTCCACTGCTGCTGCTGCTGCTGCTG 1678
1602 ATGCTCTCTCAGTCCCTCCATTTGCTTGGCTTTCCACACAGCTCAAGATGGATCCGCA 1661
1679 ATGCTCTCTAGTCCAGTGTGCTGCTGGCTCATACAA---TCATGATGGCACCAGGA 1735
1662 ACATTTAGAGATTTAGAGAGGTTACTACTGGCTTTGCGAAGACACTGCGGATGATG 1721
1736 ATATCTTAGATGATTTTAGAAGGCTTACTTTTGGCTAAGCGCAAAATACAGATGAACATG 1795
1722 CTCGGGTATGCTTTGGTGGATTAACGATACAGATAGCGGAATGGCAAAACAGAACGA 1781
1796 CAGAGTAAATGCTTTGGTGGATTTATGCTATAGATAGTGGAAATGGCTAATAGACTA 1855
1782 CGCTAGTGAATTAATACGTGGAAACATATGCTACATAGCTGCTGGTTGGCAAGCAATGT 1841
1856 CGTTGGTGGATTAATAACCTGGATAAACAGCCACATAGCACTGGTGGGAAAAGCTATGT 1915
1842 CTTCAACGAGGAGAAGCTCTACGAATTTATGACATCTCTTGAGCTGGACTAGCTTTTGG 1901
1916 CTTCTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGATGATTTGTTGG 1975
1902 TGATCTTTGGCGGTGTGATCGGCTATTCTGCGGATGATATCAACAAAGTCTCCTGTGATG 1961
1976 TTATTTTGGAGGGGTATTGGCTATTCTGCTGATGATATCAACAAATTTCTCTGGATG 2035
1962 TCCGAATTCCTGAGGAGAGATCCCAAGGACATTAAGGAAGCGATTTACTTTACCGACC 2021
2036 TTAGATAGCTGAAGAGAAACATCCCAAGACATTCGGAAGATGACTATTTTACCCAC 2095
2022 GCGGTGAATTCAGGATAGATCGGAGAGTGTCCGCGCCTGCTCAACTGCTTTATGTACA 2081
2096 AGGGAGAATTCGCTGTAGCAAAAGCAGGATCCCTCTACTTTGTTGAATGCTTTATGTATA 2155
2082 AATTAAGCTACTACAGATTCGCGGAATTTGAATTTGGACTACAGAGGTCCATCTGGATATG 2141
2156 AATGTCTACTACAGATTTGGAGAAATGCACTGGATTTTCGTACACCCCGAGTTTGG 2215
2142 ATCGCAACGTAACCGCTCATTTGGGAATAAGGACTTCGATCTGACTGACCTACCTGAGGAG 2201
2216 ACCGAACGTAATGCTGAGATTTGGAATAGGACATTAATTAACAAATTTGGAAGAG 2275

2202 CCTACACACAGAACACTGGCTTCTCGCATCTATAGGTGAAGAAGCCGATGAGTTCA 2261
2276 CTTTACATCAGAACACTGGCTTTAGGATATATAAGTAAAGCACCTGATAACAGG 2335
2262 ATAGACCATCTCTGAAGACCAAGAGAGAGATTCCTCAGCAAACTTCATTTCCAGAA 2321
2336 AGACATTAGATCAAAACCTCGAGTCAACCAATTTTCCCAAAACAGAGATTTTGTCAA 2395
2322 AGAATCTTAAGCGTCGCAAG---GGCTACATACGAAACCGACCGTGTGTTTAAGCGAA 2378
2396 AGAAGACTACAAAGGAGCGTGGCTACATTAATAAATAGCTGGTTTTTAAGAAAGCA 2455
2379 AACGAACCTTGAATAAAC 2397
2456 AGAAATATCTAAGAAGAC 2474

RESULT 4

US-10-417-375-95
; Sequence 95, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-95

Query Match 26.3%; Score 636; DB 18; Length 3046;

Best Local Similarity 58.3%; Pred. No. 1.1e-184;

Matches 1249; Conservative 0; Mismatches 800; Indels 93; Gaps 4;

118 CAGCTAATCACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177
180 CACATTTTGAAGCTTCTCATCTGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239
178 CTTGCCGCTCATCCGTTTCAGATCGATATCCAGATGATTCAGTCCGTTGTTCAACTACCG 237
240 GTTTGCTGCTCGAGATTTGAAAGTGTATCCAGATGATTCAGTCCGTTGCTGCTGCTGCT 299
238 GCCACCGCTTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
300 GACTACCAAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
298 CGCATGCTATCCGCTCGGCGAGATTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
360 AGCT 419
358 GTCCGCGGGAATCCATTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
420 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
418 CTGCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477
480 CTGCT 539
478 CAAGAGCTGTGGTCCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
540 CAAGAGCTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
538 TGGCTACATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
600 TGGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
598 CTTGAGTTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657

Db 660 CATGCTACTACCTTACTACATGTGGATCAAGGAGTAAGAGCTGGTTCCATCTGTGGGC 719
 Qy 658 GCGCGAGCGCTTTGCTTACTTCTACATGTGTCCGCTGGGTGGCTAGCTGTTCAT 717
 Db 720 AGCTAAGTGTGCTTCTTATTTTACATGGTCTGTATGGGAGGTTATGTGTTCT 779
 Qy 718 CATCAACCTGTATACCCCTGTGACGCTTCTGCTACATTTATGGGAGGTTATGCGCGC 777
 Db 780 GATCAACCTTATTTCTTCCAGCTCTGCTGCTGATGCTACAGGCGGCTTCTCTCACCG 839
 Qy 778 TCTGCTGACAGCTACAGCACTTCTATATCTTGGGAGCTGCTTCTTCCATCCAGATCC 837
 Db 840 GATCTATGTGGCTTACTGTACTTCTTACTGCTTGGGAGCTTACTTCTTACAGATCTC 899
 Qy 838 CTTGCTGGGATTCACACCGATACGACAGTGAACATGCTGGGCTGGGAGTGTGTGT 897
 Db 900 CTTTGTGGGTTTCCAGGCTGCTCTTCTCATCAGAGCACTGCGAGGCTTGTGGGCTTTGG 959
 Qy 898 GCTCTTATGGCGTGGCCACTTGGCCATTTGCGAGTCCGCTGCTGCGCAACGAGTT 957
 Db 960 TCTCTGCCAGATCCATGCTTGTGATTTACTTGGGAGTCTGCGAGAGTGAATCCACAAAT 1019
 Qy 958 CCGAAGCTGTATCTGCGGAGATGTGTGTGGGCTGTGGGCTGTGTGTGGGCTGTGTGT 1017
 Db 1020 TGAAGTTCTTTTCCGAGCGCTCATCTCTGTGTAGGCTTTGTCTCTTCTCACCGTGGAGC 1079
 Qy 1018 GGTGCTCACCATGTGGCGCTTGTGGCCCGTGGAGTGGAGCTTCTACTGCTGTGGGA 1077
 Db 1080 TCTCTCATGTGACAGAAATATCTCTCTGGAAGGCGGCTTCTACTGCTGCTGGA 1139
 Qy 1078 TACTGGCTACGCCAAGATCCATCTCCATCATCTGCTGCTGGGCTGTGGGAGCATCACCCAC 1137
 Db 1140 TCCCTCTTATGCTAAGAACACATCCCATCATCTGCTGTGTGCTGTGAGCATCACCCAC 1199
 Qy 1138 CACTTGGTCTCTCTTCTTCTTCTGATCTGCACTCTCTGCTGTGGGCTTCCAGTGGAGT 1197
 Db 1200 AACCTGGTCTCATCTATTTTACCTGAGCTCTCTCTCTCATGTTCAGTTGGGCT 1259
 Qy 1198 GTGGTACTGTACAGCAGATCAACGAGAGCGCTTTCGTGGTGTCTAGCCCATCAG 1257
 Db 1260 CTATTTACTGCTTTAGCAACCTGTCTGATGCGCCGATTTTATCATCATGTATGGTGTAC 1319
 Qy 1258 TGCGTCTTACTTCTGCTGT 1317
 Db 1320 CAGCATGTCTTTTCTGCTGT 1379
 Qy 1318 CATGCTGGCGAGTGGCTTTTGGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1377
 Db 1380 CATTTCTCTGCGATTTGGAGTCTCCAGGTGCTG----- 1413
 Qy 1378 TAAGCAATGGGCACAGCCATTAAGCGCAGCCAGCTGAGTGAAGCTGAGGATTCAT 1437
 Db 1414 -----TCCAC 1418
 Qy 1438 TGAGAGAGAGCTGTACACAGAGCTGCCAGCTGAAGCATGTGTAACCATGTATGC 1497
 Db 1419 ATATGAAGAATCTGACATAAGT 1478
 Qy 1498 CCAGCAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
 Db 1479 CACCTACCTATT-----AAGATGAAGTGGCAAGTGGATGATGATGATGATGATGATGAT 1532
 Qy 1558 CTTGT 1617
 Db 1533 CTTTCTCATCACTTACACCTTTTCTTCAACCTGGGTGACGAGGCTTCTTCTTCTTCTTCT 1592
 Qy 1618 CTCATTGTCTTGTGCTTTTCCACACACAGTCAAGATGATCCGCAACATTTTACAGATTT 1677
 Db 1593 GTCCATTGTACTATCTGCC-----GTGGTGGGATGGCAGTAGGATCATATTTGATGACTT 1649
 Qy 1678 CAGAGGCTTACTTACTGCTTTTCCGAGACATCTGCGATGATGCTCGGCTTATGTCTGTG 1737

Db 1650 CCGAGAGCATATTATTGGCTTCTGCTATTAATCTCCAGAGGATGCGAAGTCTGATGCTGT 1709
 Qy 1738 GTGGGATTCAGGATACAGATAGCGGGAATGGGAAACAGAAACAGAGCTAGTGTGATATAA 1797
 Db 1710 GTGGGATTCAGGATACAGATAGCGGGAATGGGAAACAGAAACAGAGCTAGTGTGATATAA 1769
 Qy 1798 TACGTGGAAACAATAGTCTACATAGCGCTGTGGTGGGAGGCAATGCTTCAACCGAGAGAA 1857
 Db 1770 CACATGGAAATATATCCCATATTTCTCGAGTAGGCGAGGCAATGGCGTCCACAGAGGAAA 1829
 Qy 1858 GTCTTACGAAATATGACATCTCTGACGTGAGTCTAGCTTTTGGTGTATCTTTGGCGGTGT 1917
 Db 1830 AGCTATGAGATCATAGGAGCTCGATGTAGCTATGTCTGTCTATTTTGGAGGCT 1889
 Qy 1918 GATCGGCTATTTTGGCGATGATATCAACAAGTCTCTGTGTGATGCTCGAATTTGCTGAGGG 1977
 Db 1890 CACTGGGATTTCTCTGATGATATCAACAAGTCTTCTTGGATGCTCGGATTTGGAGGAG 1949
 Qy 1978 ---AGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCAGCGCGTGAATTCAG 2034
 Db 1950 CACAGATACAGCAACATATCAAGGAGATGACTATTAATCTCCAATCGGAGTTCCG 2009
 Qy 2035 GGTAGATGCCAGAGTGTCTCGGCGCTGCTCACTGCTTATGTACAAATTAAGTACTA 2094
 Db 2010 TGTGGACGCTGAAGTCTTCCAGTGTCTCACTGCTCATGTGACAGATGTGTACTA 2069
 Qy 2095 CAGATTCGGGGAATTTGAAGTTGGACTACAGAGTCTCATCTGGATATGATCGCACAGTAA 2154
 Db 2070 TCGTTTGGACAGGTTTACACAGAGCAAGCGTCTCCAGGCTTTGACCGTTCGGA 2129
 Qy 2155 CCGCTGTATTTGGGAAATAGGACTTGTGATCTGACCTACTGAGGAGGCTTACACACAGA 2214
 Db 2130 TGTGAGATTTGGAAATTAAGACTTTGAGCTTGTCTCTGGAGGAGCATATACACAGA 2189
 Qy 2215 ACATGGCTTGTTCGCACTATAGGTTGAAGAGCGCGATGA 2256
 Db 2190 ACATGGCTTGTTCGCACTATAGGTTGAAGAGCGCTGATGA 2231

RESULT 5

US-10-128-714-7139
 ; Sequence 7139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wendi
 ; APPLICANT: Tishkoff, Daniel
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Broshkin, Alexey M
 ; APPLICANT: Lemieux, Sebastien M
 ; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 ; TITLE OF INVENTION: Methods of Use
 ; FILE REFERENCE: 10182-018-999
 ; CURRENT APPLICATION NUMBER: US/10/128,714
 ; CURRENT FILING DATE: 2002-04-23
 ; PRIOR APPLICATION NUMBER: US 60/285,697
 ; PRIOR FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: US 60/287,066
 ; PRIOR FILING DATE: 2001-04-27
 ; PRIOR APPLICATION NUMBER: US 60/295,890
 ; PRIOR FILING DATE: 2001-06-05
 ; PRIOR APPLICATION NUMBER: US 60/303,899
 ; PRIOR FILING DATE: 2001-07-09
 ; PRIOR APPLICATION NUMBER: US 60/316,362
 ; PRIOR FILING DATE: 2001-08-31
 ; NUMBER OF SEQ ID NOS: 8603
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 7139
 ; LENGTH: 2232
 ; TYPE: DNA
 ; ORGANISM: Aspergillus fumigatus
 US-10-128-714-7139

| | | | | | |
|-----------------------|------|--|---------------------|------------|--------------|
| Query Match | | 26.3%; | Score 635.4; | DB 15; | Length 2232; |
| Best Local Similarity | | 58.9%; | Pred. No. 1.4e-184; | Indels 75; | Gaps 6; |
| Matches 1248; | | Conservative 0; | Mismatches 196; | | |
| QY | 137 | ATCCGTGTAATCGCTGCTGCGCGGATTTTCCTCTCGCCCTTTTCGCGCTCATCGGTTTC | 196 | | |
| DB | 76 | ATCCGTGTACCATTCGACGACGACAGTTCGCCAGTCGACTCTTCAGCGTTATCGGCTTC | 135 | | |
| QY | 197 | GAGTCGATATCCAGTATTCGATCCGTTGATCCGTTTCAACTACCGGCGACCGCTACATGCTG | 256 | | |
| DB | 136 | GAGATATCATCCAGAGTTGACCCGTTGTTCACTTCGAGGACAAATATCTAGTA | 195 | | |
| QY | 257 | CAGATATGTTGTAACAATCTCTCAACTGTTTCGACGAGCGCATGGTATCCGCTCGGC | 316 | | |
| DB | 196 | CAGATATGTTTCTATAGCTTTTGGGATTTGTTGATGACGGAACATGGCATCTCTCGGA | 255 | | |
| QY | 317 | AGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTAGCTCCGCGGAATCCATTTG | 376 | | |
| DB | 256 | CGTGTACCGGTGGGACGTTATATCCCGGCTCTATGTTGACGAGCGGCTGATCTACCAT | 315 | | |
| QY | 377 | CTGCTGACGTAATCAACATACCGGTTCATATTCGTGACATCTGCTGTTCTCTGGGCGG | 436 | | |
| DB | 316 | ATCTTGGCATTCCTTACTATATCCCGCTGATTTGCAACATCTGCTCTACTGGGCGCA | 375 | | |
| QY | 437 | ATCTTCAGTGGCTGACCTCATCTCCACTACCTGCTGACCAAGGAGCTGCTCGCG | 496 | | |
| DB | 376 | GGATTTCTCGGCTGACTGCAATGGCAATGATTTGCTGATCCGAGATGTTCTCTTCG | 435 | | |
| QY | 497 | GGC---GGCGGCTCTTCGCGCCAGCTTCATCCCATTCGCTGCTGCTGCTACATGATGG | 553 | | |
| DB | 436 | CCATCTCGAGGTCTTCTTGCAGCAGCTTTCATGGGAATCGCCCTGTTTACATCTCCCGA | 495 | | |
| QY | 554 | TCGGTGGCTGATCGTACGATAACGAGGAGTTCGCATATTCGCCCTCGAGTTCACTAC | 613 | | |
| DB | 496 | TCAGTGTCTGAGAGTACGATAACGAGGATTCGCATCTTCTGCTGTTGTTCAATTC | 555 | | |
| QY | 614 | TTCTGTGGTGGCTGCTGAGTGAAGTCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTG | 673 | | |
| DB | 556 | TTTCTATGATCAAGGCTGTCAAAAATGGTCTATCATGTGGGAGGCTGACCGCACTA | 615 | | |
| QY | 674 | TCCTACTTCTACATGTTGCTGCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 733 | | |
| DB | 616 | TTCTACGGCTACATGGTGTGCGATGGGTTGGTATGTTCTTCAATACGAACTGATCCCC | 675 | | |
| QY | 734 | CTGCAAGTCTTCTGCTCATATATGGGAGGTAATCGCGCTCTGCTGACCACTAC | 793 | | |
| DB | 676 | CTGCAAGTCTTCTGCTCATATGGGAGGTAATCGCGCTCTGCTGACCACTGATAGCTAT | 735 | | |
| QY | 794 | AGCAGCTTCTACATCTGCGAGTCTGCTTCTCCATGAGATCCCTTCGTGGGATCCAA | 853 | | |
| DB | 736 | ACCATATGATGCGCTGGGAGCTTTGGCTAGCATGCAATTCCTTCGTGCGGATTTTG | 795 | | |
| QY | 854 | CCGATACGCAAGTGAACACATGGCTGCTGCGGAGTGTGTTGCTCTTATGGCGGTG | 913 | | |
| DB | 796 | CCTATCCGAAACAGCGACCAATGTCCGCTTGGGTGCTTCGCGCTGCTTACGTTGTG | 855 | | |
| QY | 914 | GCCACTTGGCGCATTTTCAGTCTGCTGCTGCGGACGATTCGCGAGCTGTTTATC | 973 | | |
| DB | 856 | GCTTTCGCGGATTTGCTCGAGCTTCGTTCCAGAGAGAGTTCGAGAGCTTCTGACC | 915 | | |
| QY | 974 | GTCCGCGGATTTGCTGGGCTTGGCTCTTTGTCGCGCTGCTGCTGCTGCTGCTGCTG | 1033 | | |
| DB | 916 | GCATGATCTTCTACCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT | 975 | | |
| QY | 1034 | GGCGTGTGGCGCTGGAGTGGAGCTTCTACTGCTGCTGGGATCTGGCTAGCCCAAG | 1093 | | |
| DB | 976 | GGAGTATGCTTCTGGAGCGGCGGATCTTACTCTTGTGGGAGACTGGCTATGCCCAA | 1035 | | |
| QY | 1094 | ATCCAGATTTCCATTTGATTCGTTGCGGAGATCAGCCCACTTGGTTCTCGTTTC | 1153 | | |
| DB | 1036 | ATCCAGATTTCCATTTGATTCGTTGCGGAGATCAGCCCACTTGGTTCTCGTTTC | 1095 | | |
| QY | 1154 | TTCTTTGATCTGCAATCTGCTGCTGCTGCTTCCAGTGGGAGTGGTACTGATCAAG | 1213 | | |
| DB | 1096 | TTCTTCGATCTGAACTTCTCTGATCTGCTTTTCCCGCAGGTGCTTACATGTTCTCGCT | 1155 | | |
| QY | 1214 | CAGATCAACGACGAGCGGTTTTCGTGCTGTAGCCCATAGTGGGTTTACTTTCGCT | 1273 | | |
| DB | 1156 | GACCTCAAGGACGAGCATGTCTTCTGATTAATCTACTCGGTCTCTGCGAGCTACTTTCGC | 1215 | | |
| QY | 1274 | GCTGTGATGGTGGCTTTGATTTGACCTCAACCGGTGGTGTGATGCTGGCGGAGTG | 1333 | | |
| DB | 1216 | GCTGTATGTTCGACATAATGCTGACCTTGACCCCTTATTTGTGTGTGTGCGCTGCTCTG | 1275 | | |
| QY | 1334 | GCTTTTTCGGAAGTGTGATGTTCTCTCAAGAGGATTCGTTCAAGCAATGGGACACA | 1393 | | |
| DB | 1276 | GCGTGTGCTCCATTTCTCGACACCTATATG----- | 1305 | | |
| QY | 1394 | GCCATAGCGCAGCACCGAAGTGAAGCTGAGGATTCATTCAGAGAAGAGCGCTG | 1453 | | |
| DB | 1306 | GCGACTACCTCCCGACACCGAGCTCTGAAGC----- | 1338 | | |
| QY | 1454 | TACGACAAAGGTGCAAGCTGAGGATCTGTAAGCATGATCCCGACGAGGATATCGGC | 1513 | | |
| DB | 1398 | AAAAAGCAATGAAGACTCGTCTTCAACCTCTTCGCTCAAGTTCGGAAGCCCAATGTGGA | 1398 | | |
| QY | 1514 | GTCACTCTCAACCTG---AAGAGTATTTGTTTGGCCGTTCTAATGCTGTTGATGATG | 1570 | | |
| DB | 1399 | ATCACTTCCCATGTTTCTAAGATTATAGTCAGCGGCTCTGTTCTGCTACCTGCTCTG | 1458 | | |
| QY | 1571 | TTGCTGTCTCACTGCAAGTGGTGAACGCAATGCTTCTCACTGCTGCTGCTGCTGCTG | 1630 | | |
| DB | 1459 | TTTGTGGCACTGCACTGGGTTTACATCGAATGCACTCTTCTCTCTGCTGCTGCTG | 1518 | | |
| QY | 1631 | GCTTTTCAACACAGTGAAGTGAATCCCGCAACATTTAGACGATTTACAGAGAGGTTAC | 1690 | | |
| DB | 1519 | GCTAGTCGATGCTCT---GACGGAAGCAATACATCATTCACGATTTCTGAGGCTTAC | 1575 | | |
| QY | 1691 | TACTGCTTTTCGAGAACACTGCGGATGATGCTGCGGTATGCTGCTGCTGCTGCTGCTG | 1750 | | |
| DB | 1576 | TACTGCTTCTGATGATCTCTCAGACGCAAAATCATGTGCTGCTGCTGCTGCTGCTG | 1635 | | |
| QY | 1751 | TACGATAGCGGGAATGGGAAACAGAACGACGCTAGTGATTAATATAGTGGAAAT | 1810 | | |
| DB | 1636 | TATCAATCGTGTGATGCGGACCGCCCAACCTTGGTTGACAAACACCTCGGAACAC | 1695 | | |
| QY | 1811 | AGTCAATAGCGCTGTTGGCAAGCAATGTCTTCAACGAGGAGAGTCTTACGAAAT | 1870 | | |
| DB | 1696 | ACCCATATTTGCTAGGTTGGTAAAGCGATGAGCTCAAGGAGAGTCACTACCCCATC | 1755 | | |
| QY | 1871 | ATGATCTCTTACGCTGAGTACGTTTGTGATCTTTGCGGCTGCTGCTGCTGCTTCT | 1930 | | |
| DB | 1756 | CTCGCCAGCATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTCT | 1815 | | |
| QY | 1931 | GGCGATGATCAACAGTTCCTGTGATGCTGCGAATTCCTGAGGAGAGCATCCCAAG | 1990 | | |
| DB | 1816 | GGCGATGATCAACAGTTCCTGTGATGCTGCGAATTCCTGAGGAGAGCATCCCAAG | 1875 | | |
| QY | 1991 | GACATTAAGGAAAGCATTTACTTTACGACCGCGGTGAATTCAGGGTAGATGCCAAGGT | 2050 | | |
| DB | 1876 | GAGGTAAAGAGCGGACTCTTTTACTGACGCGGTGAATATCGTGTGAGATGAGCG | 1935 | | |
| QY | 2051 | GCTCGGCGCTGCTCAACTGCTTATGTAACAATTAAGCTACTACAGATTCGGGGAATTG | 2110 | | |
| DB | 1936 | ACCCCAACTATGCGCAACAGCTTGTGTATATAAATGCTTTATTACA-----ATTTC | 1986 | | |
| QY | 2111 | AAGTTGGAATACAGAGGTCTCATCTGGATATGATCGCACAGTAAACGCGCTCATTTGGG | 2170 | | |
| DB | 1987 | ACTCTCTCTTCCGCTCGGCGCAAGCTGCGACGCGCTCGTGGGTCAAACTTCCACA | 2046 | | |
| QY | 2171 | AAGGACTTCGATCTGACCTACTCGGAGGCGCTTACACCAAGAACTCTGCTGCTGCTG | 2230 | | |
| DB | 2047 | GAGGCGCTCTGCTCTTACATCTCAAGAGAGCTTTTCGAGCGAGAACTGGATCTTCTG | 2106 | | |
| QY | 2231 | ATCTATAGGTTGAAGAGC 2249 | | | |

Db 2107 ATCTACAGGTCAGGATC 2125

RESULT 6
US-10-171-581-112
; Sequence 112, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mac Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 112
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L38961
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-112

Query Match 26.2%; Score 632.8; DB 15; Length 2472;
Best Local Similarity 58.2%; Pred. No. 9.4e-184;
Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;

QY 118 CAGCTAATCACCTTCGCCATCTGCTAATCGCTCGCTGGCGGATTTCTCTCGCT 177
DB 151 CACACTTTGAGCTTCTCAATCTGCAATGGCTGCTGATTTATCTCTCTCCACTCGTCT 210
QY 178 CTTCGCGTCAATCGCTTCGAGTCGATTTCCATGAGTTTGAATCCGCTGCTCACTACCG 237
DB 211 GTTTGCTCTCTGAGATTTGAAAGTGTATCCATGAGTTGATCCGCTACTTTAATATCG 270
QY 238 GGCACCGCTACATGCTGAGATGTTGGTACAACTTCTCACTGCTTCGAGGCG 297
DB 271 GACTACAGGTTCTGCTGAGGAGGTTTATAAATTTCCATCTGTTGATGACCG 330
QY 298 CGCATGATTCGCTCGGAGGATTTGGCGGCTACCGCTCTATCCCGGCTGATGATAC 357
DB 331 AGCTGTGATCCCTTTGGGACGGAATCATTTGAGGAAACAATTTACCCAGGTTTAATGATC 390
QY 358 GTCCGCGGGAATCCATTGGCTGCTGACGCTACTCAACATACCGTCCATATTTCTGATC 417
DB 391 CTCTGCTGCAATCTACCATGATCTCCATTTTTCACATCACATCGACATTCGGAATGT 450
QY 418 CTGCGTCTCTCGCGCGCATTTTCAGTGGCTCGACCTCCATCTCCACTACCTGCTGAC 477
DB 451 CTGTGTCTCTCGCGCGCT 510
QY 478 CAAGGAGCTGTGTTCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537
DB 511 CAAGAGCTCAAGATGACGAGGCTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 570
QY 538 TGGCTACATCAGTAGTGTGCTGCTGATCGTACGATGAGGAGGATTCGATATTCGC 597
DB 571 TGGATATATCTCCGATCTGTGGCTGGCTCTATGATATGATGAGGATTCGATCTTTTG 630
QY 598 CTTGAGTCTCACCTACT 657
DB 631 CATGCTACTCACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 690
QY 658 GCGCGAGCGCTTTGCTCTACTTTCTACATGAGTGTCCGCTCGGCTGCTGCTGCTGCT 717
DB 691 AGCTAAGTGTGCT 750
QY 718 CATCAACCTGATACCCCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 777

Db 751 GATCAACTTAATCTCTCTCCAGTCTCTGCTGCTGATGCTCACAGCGGCTTTCTCTCACCG 810
QY 778 TCTGCTGACCACTACAGCACTTCTACATCTCTGGACTGCTGTTCTTCCATGACAGATCCC 837
DB 811 GATCTATGTGGCTACTGCTACTGTTTACTGCTGGTACTATATCTTTCTAGGCAGATCTC 870
QY 838 CTTCTGTTGATTTCCAACTGATACGACAGTGAACACATGCTGCTGGGAGTGTGT 897
DB 871 CTTTGTGGTTTCCAGCTGCTCTTTTCAAGACATGACAGGTTTGGGCTTTGG 930
QY 898 GCTCTTATATGCGGCTGGCCACCTTTCGCAATTTGAGTCTCTGCTGCGCAACAGATTT 957
DB 931 TCTCTCCAGATCTCATGCTTGTGATTTACCTGCGACAGTGAATCCACAACAAT 990
QY 958 CCGGAGCTGTTTCACTGCTGCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
DB 991 TGAAGTTCTTTTCCGAGGCTCATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1050
QY 1018 GGTGCTCACTGCTGCGGCTTTGGCCCTGCTGAGTGAACGCTTCTACTGCTGCTGCTG 1077
DB 1051 TCTCTCTCATGCTGACAGGAAATATCTCTCTGCGAGCGGCTTTCTACTCTCTCTGGA 1110
QY 1078 TACTGCTACGCAAGATCCCACTTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
DB 1111 TCGCTTATGCTAAGAACATCTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
QY 1138 CACTGCTTCTGCTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
DB 1171 AACCTGCTCTCATATTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1230
QY 1198 GTGGTACTGCTAAGCAGATCAACGACGAGCGGCTTTTCTGCTGCTGCTGCTGCTGCT 1257
DB 1231 CTATTAAGCTTTTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1290
QY 1258 TGGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
DB 1291 CAGCATGATCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1350
QY 1318 CATGCTGCGGAGTGGCTTTTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1377
DB 1351 CATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1384
QY 1378 TAAGCAATGAGCAGACCAATAGCGCAGCCACCAAGTGAATGAGCTGAGGATTCAT 1437
DB 1385 -----TCCAC 1389
QY 1438 TGAGAAGAGAGCTGTACGACAGGCTGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1497
DB 1390 ATACATGAGATCTGAGATAGTCCGCCAGACAGAGAGCAAGCAAGATTC 1449
QY 1498 CAGCAGGATACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1557
DB 1450 CACCTACCTATTAAGATTT-----GAAGTGGCAAGTGGGATGATGCTGCTGCTGCT 1503
QY 1558 GCTGTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1617
DB 1504 CTTTCTCATCACTACACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1563
QY 1618 CTTCAATGCTTGGCTTTTCCACAAAGTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTG 1677
DB 1564 GTCCATTTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
QY 1678 CAGAGAGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737
DB 1621 CCGAGAGCATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
QY 1738 GTGGGATTAAGATACAGATAGCGGAAATGCGAAACAGAACAGACGCTGCTGCTGCTG 1797
DB 1681 GTGGGATTAAGATACAGATAGCGGAAATGCGAAACAGAACAGACGCTGCTGCTGCTG 1740
QY 1798 TACGTGGAAACATAGTACATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1857
DB 1741 CACATGGAATATACCATATTTTCTCGAGTAGGCGAGCAATGGCTGCTGCTGCTGCTG 1800

QY 1438 TGAGAGAGAGAGCTGTATGACAGAGGCTGCAAGCTGAAGCATCTGACTAAGCATGATGC 1497
 Db 1390 ATACATGAAGAAATCTGACATAAGTCCGCCAGAGCAAGAGAGAGCAAGCAAGCAAGATTC 1449
 QY 1498 CCAGCAGGATAGTGGCGTCAAGCTCCACCTGAAGAGTATTTGTTTGGCGGTTCTAAT 1557
 Db 1450 CACTACCTATTAGATTT-----GAAGTGGCAAGTGGAGATGATCTGTCATGGCTTT 1503
 QY 1558 GCTGTGTAGATGTTGCTGCTCACTGCAAGTGGGAGACAGCAAGTGGCTTATCCAGATCC 1617
 Db 1504 CTTTCTCATCACTACACCTTTTCAATTCACCTGGGTGACCAAGTGGAGGCTTCTTCTCC 1563
 QY 1618 CTCATTTCTTGTGCTTCCACACAGTCAAGATGATCCGCAACATTTTGAAGATTT 1677
 Db 1564 GTCCATTTGATATCTGCTCC-----GTGGTGGGATGGCAGTATGATTTGATGACTT 1620
 QY 1678 CAGAGAGGCTTACTACTGCTTTCGCAAGACACTGGCGATGATGCTGGGCTTATGCTTG 1737
 Db 1621 CCGAGAGCATATTTATTTGGCTTCTCATATAATCTCCAGAGGATGCGAAGGTCAATGCTCG 1680
 QY 1738 GTGGGATTTACGGATACAGATAGCGGAATGGCAACAGAGCAAGCTAGTGGATAATAA 1797
 Db 1681 GTGGGATTTAGCTATCAGATACAGATATGGCAACAGAGCAATTTTATGAGCAATAA 1740
 QY 1798 TACGTGGAAATAGTACATAGCTGCTTTCGCAAGCAAGTCTTCAACCGAGAGAA 1857
 Db 1741 CACATGGAAATATACCCATATTTCTCGAGTAGGGCAGCAATGGCTCCACAGAGGAAA 1800
 QY 1858 GTCCCTACCAATTTATGACATCTCTGAGTGGGATGATGTTTGGTATCTTTTGGGGGTG 1917
 Db 1801 AGCTATGAGATCATGAGGAGCTGATGTCAGTATGCTGCTGCTATTTTGGAGGCT 1860
 QY 1918 GATCGGCTATTTGCGGATGATATCAACAGTTCCTGTTGGATGGTCCGAATTTGCTGAGG 1977
 Db 1861 CACTGGGTATTTCTCTGATGATATCAACAGTTCCTTGGATGGTCCGAATTTGAGGGAG 1920
 QY 1978 ---AGAGCATCCCAAGGATATTAAGGAAAGCGATTTACTTTACCGACCGGGTGAATTCAG 2034
 Db 1921 CACAGATACAGCAACATATCAAGAGAGATGACTATTATCTCCACTGGGAGTTCG 1980
 QY 2035 GTAGATCGCAAGGTTCTCGGCTGCTCACTGCTTATGATGCTTATGATAAATTAAGTACTA 2094
 Db 1981 TGTGACCGGTGAAGGTTCTCGAGTGTCTCACTGCTTCACTGCTCATGATCAAGATGTTACTA 2040
 QY 2095 CAGATTTGGGAAATGAAGTTGGACTACAGAGGTCCATCTGGATATGATCGCACACGTA 2154
 Db 2041 TCGCTTTGGACAGGTTTACACAGAGCAAGCGTCTCCAGGCTTTGACCGGTGTCGAA 2100
 QY 2155 GCGGTGATTTGGGAATAGGACTTCGATCTGACCTACCTGGAGGAGGCTTACACACAGA 2214
 Db 2101 TGCTGAGATTTGGGAATAAGACTTTGAGCTTGATGCTGGAGGAGGCTATACACAGA 2160
 QY 2215 ACACCTGGCTGCTGCTGCTATAGGCTGAAGAGCGCGCATGA 2256
 Db 2161 ACATTTGGCTGCTCAGGATATACAGGTTAAGGACCTGGATAA 2202

RESULT 8

US-10-172-118-742
 ; Sequence 742, Application US/10172118
 ; Publication No. US20030224374A1

GENERAL INFORMATION:

; APPLICANT: Dai, Hongyue
 ; APPLICANT: He, Yudong
 ; APPLICANT: Linsley, Peter
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Roberts, Chris
 ; APPLICANT: Van 't Veer, Laura
 ; APPLICANT: Van de Vijver, Marc
 ; APPLICANT: Bernards, Rene
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
 ; FILE REFERENCE: 9301-175-999
 ; CURRENT APPLICATION NUMBER: US/10/172.118

; CURRENT FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: 60/380,770
 ; PRIOR FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 2699
 ; SEQ ID NO 742
 ; LENGTH: 2472
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: NM_002219
 ; DATABASE ENTRY DATE: 2001-06-18
 ; US-10-172-118-742

Query Match 26.2%; Score 632.8; DB 15; Length 2472;
 Best Local Similarity 58.2%; Pred. No. 9.4e-184; Mismatches 802; Indels 93; Gaps 4;
 Matches 1247; Conservative 0;

QY 118 CAGCCTAATCACTTGGCCATCTGCTAATCGCTGGCTGGCGGATTTCCCTCTCCCT 177
 Db 151 CACACCTTTGAAGCTTCTCAATCTGCTAATGGCTGCTGATTAATCTCTCCACTCGTCT 210
 QY 178 CTTGCGCGTCATCCGTTTCGGAGTCGATTATCCATGAGTTTGCATCCGTGGTTCAACTACCG 237
 Db 211 GTTGTGCTTCGAGATTGAAGTTGATTCATGAGTTTGCATCCGTACTTTAATATCG 270
 QY 238 GGCACCGCCTACATGGTGCAGAAATGGTTGGTACAACTTCTCACTGGTTGCAGAGAG 297
 Db 271 GACTACAGGTTCTCGCTGAGGAGGGTTTTTATAAATTTCCATTAATCTTGTATGACCG 330
 QY 298 CGCATGATATCGCTCGGCGAGGATTTGGCGGATACCGTCTATCCCGCTGATGATTAAC 357
 Db 331 AGCTGGTACCTTTGGGACGATCAATTTGAGAGCAATTTACCAGGTTTATGATCAC 390
 QY 358 GTCGCGCGGAATCCATTTGGCTGCTGACGATCACTCAACATACCGGTCATATTCGTGACAT 417
 Db 391 CTCTGTGCAATCTACCATGCTACTCCATTTTTCACATCAACATCCATCCGATTCGGAATGT 450
 QY 418 CTGCGTTCCTGGCGCGGATCTTCCAGTGGCTGACCTCCATCTCCACCTACCTGCTGAC 477
 Db 451 CTGTGTGCTTGGCGGCT 510
 QY 478 CAAGAGCTGTGGTCCGCGGCGCGGCTCTCTGCGCGGCGAGCTTCACTGCGCATCTGCGCC 537
 Db 511 CAAAGAGCTCAAGGATGAGGCGCTGGGCTTCTTGTCTGCTGCCATGATTCGTGATGTC 570
 QY 538 TGGCTACATCAGTAGTGGTGGGCTGAGTCTGATGATTAAGAGGCGATTTGCCATATTCGC 597
 Db 571 TGGATATCTCCGATCTGTGGCTGGCTCTCTATGATTAAGAAGGATTTGCCATCTTTG 630
 QY 598 CTTGCAAGTTCACTTCTCTGCTGGGCTGGCTCAAGTGAAGTGGATCCGTTTCTGGTC 657
 Db 631 CATGCTACTCACTACTACATGTTGATCAAGGCGAGTAAAGACTGGTTCCATCTGTGTGGGC 690
 QY 658 GGCAGCGAGCGGTTTGTCT 717
 Db 691 AGCTAAGTGGCTTGTCTTCT 750
 QY 718 CATCAACCTGATACCTCTGAGCTTCTGATCTGCTCATTTATGAGGAGGCTACTCTGCCGCG 777
 Db 751 GATCAACTTAATTCCTCTCCAGCTCTGCTGATGCTCAAGGCGGTTTCTCTCCACCG 810
 QY 778 TCTGCTGACAGCTACAGCACTCTTACATCTGGGAGCTGCTTCTCCATGAGAGATCCC 837
 Db 811 GATCATGTGGCTACTGTACTGTTTACTGCTGGTACTATATCTTCTAGGAGAGATCTC 870
 QY 838 CTTGCTGGGATTTCCAAACCGATACGCAACAGTGAACATGCTGCTGGCTGGAGTGTGTGT 897
 Db 871 CTTTGTGGGTTTCCAGCTGCTCTTTCATCAGAGCACATGCGAGGTTTGGGGTCTTTGG 930
 QY 898 GCTCTCTTATGGCGGTGGCCACCTTTCGCGCATTTTGCAGTCCGCTGCTGCGGCAACAGAT 957
 Db 931 TCTCTGCCAGATCCATGCTTGTGTTGATCTGCGGAGCAAGTTGAATCCACAACTT 990

QY 958 CCGAAGCTTTCATCGTCGGGGATGCTGGTGGCGTGGCGTCTTTGGCGCGT 1017
Db |||||
QY 991 TGAAGTCTCTTCGGAGGCTCATCTCTCTGTAGGCTTTGTCTCTCACCGGGAGC 1050
Db |||||
QY 1018 GTGTCTCACCATGCTGGGCTTGTGGCCCGTGGAGTGGAGCTTCTACTCGTGGGA 1077
Db |||||
QY 1051 TCTCTCTATGCTGACAGGAAATATCTCTTGGACGGGGCTTCTACTCAGTGTGGA 1110
Db |||||
QY 1078 TACTGCTACGCAAGATCCATTCATCCCATCATCTCCGTGTGGAGCATCAGCCAC 1137
Db |||||
QY 1111 TCCCTCTTATGCTAAGAACCAATCCCATCATCTCTCTGTCTGAGCATCAGCCAC 1170
QY 1138 CACTGGTCTGCTCTCTCTTGTATCTGCATCTGGTGTGGCTTCCAGTGGAGT 1197
Db |||||
QY 1171 AACCTGCTCTCATCTATTTGACCTGAGCTCTCTCTCTCATGTTTCCAGTGGCT 1230
QY 1198 GTGTACTCATCAAGACATCAACGACGAGCGCTTTTCGTGGTGTGTACGCCATCAG 1257
Db |||||
QY 1231 CTATTACTGCTTTAGCAACCTGCTGATGCCGGATTTTATCATCATGTATGTTGAC 1290
QY 1258 TCGGCTTACTTCTGCTGTGTGATGCTGTGTGATGTGACCTCAGCGGCTGTG 1317
Db |||||
QY 1291 CAGCATGTAATTTTCTGCTGTATGCTGTGCTGTATGCTGTGCTGTGCTGTGAG 1350
QY 1318 CATGCTGCGGAGTGGCTTTTCGGGACTGTGTGATGTGTCTGTCAAGAGGATTCGTC 1377
Db |||||
QY 1351 CATCTCTCTGCAITGGAGTCTCCAGGTGCTG----- 1384
QY 1378 TAAGCGAATGGCAGACCATTAAGCGCAGCCACCGAAGTGAAGTGAAGTTCAT 1437
Db |||||
QY 1385 -----TCCAC 1389
QY 1438 TGAGAAGAACGCTGTACGACAGGCTGGCAAGCTGAAGCATCGTACTAAGCATATGC 1497
Db |||||
QY 1498 CCAGCAGGATCTGGGCTCAGCTCCCACTGAGGATTTCTTTTGGCGTCTAAT 1557
Db |||||
QY 1504 CTTCTCTCATCCTACCTCTTCAITCAACCTGGGTGACAGTGAAGGCTACTCTTCTCC 1563
QY 1618 CTTCTCTCTGCTTCTTCCACAGTCAGATCGATCCCGCAACATTTTAGACGATTT 1677
Db |||||
QY 1564 GTCCATTTACTATCTGCCC-----GTGGTGGGATGGCAGTGAATCATTTGTAGACT 1620
QY 1678 CAGAGAGCTTACTACTGCTTTTCGAGAACACTGCGCATGATCTCGCGTATGCTTG 1737
Db |||||
QY 1621 CCGAGAAGCATATTTATGCTTCTGTCATATACTCCAGAGATGCAAGGTCTGCTG 1680
QY 1738 GTGGGATACGATACGAGGATGCGAAGACAGACGAGCTAGTGGATATAA 1797
Db |||||
QY 1681 GTGGGATATGCTATCAGATTACGCTATGCAACCGAACAATTTTAGTGGCAATAA 1740
QY 1798 TAGGTGGAACAATAGTCAATAGCGCTGTTGGCAAGGCAATGTCTTCAACCGAGGAA 1857
Db |||||
QY 1741 CACATGGAATATACCAATTTCTCGAGTAGGCGAGCAATGGCTCCACAGAGGAAA 1800
QY 1858 GTCTACGAAATATGACATCTTTGAGTGAATGCTGTTTGGTGTCTTGGCGGT 1917
Db |||||
QY 1801 AGCTATGATATCAGAGGAGCTCGATGTCAGCTATGCTGTGCTCAATTTTGGAGCCT 1860
QY 1918 GATCGGCTATTTGGCGATGATATCAACAAGTCTCTGTGGATGTGCGAATGTGAGGG 1977
Db |||||
QY 1861 CACTGGGTATCTCTGATGATATCAACAAGTCTCTTGGATGTGCGGATTTGGAGGAG 1920
QY 1978 ---AGAGCATCCAGGACATTAAGGAGGATTTACTTTACCGACCGGCTGATTCAG 2034
Db |||||
QY 1921 CACAGATACGGAACAATATCAAGAGAAATGACTATTTATCTTCACTGGGAGTCCG 1980
QY 2035 GGTAGATGCGAGGTGCTCGCGCCCTGCTCAACTGCTTATGATCAAAATTAAGCTACTA 2094

Db 1981 TGTGACCGTGAAGGTTCTCCAGTGTCTCAACTGCTCATGTACAAGATGTGTACTA 2040
QY 2095 CAGATTCGGGAAATTGAGTTGGACTACAGAGTCCATCTGGATATGATCGCACGTAA 2154
Db 2041 TCGCTTTGACAGGTTTACAGAGAGCAAGCGTCTCCAGGCTTTGACCGTTCGAAA 2100
QY 2155 CCGCTCATTTGGGAATAGGATTCGATCTGACCTACCTGGAGAGGCTTACACACAGA 2214
Db 2101 TGCTGAGATTGGGAATAAGACTTTGAGCTTTGATGTCTCGAGGAAGGCTTATACCAGA 2160
QY 2215 ACACCTGGCTTGTTCGCATCTATAGGTTGAAGAGCCCGCATGA 2256
Db 2161 ACATTTGGCTGTGAGATATACAGGTAAGGACCTGGATAA 2202

RESULT 9

US-10-342-887-742
; Sequence 742, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Mao, Mao
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, René
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-742

Query Match 26.2%; Score 632.8; DB 16; Length 2472;
Best Local Similarity 58.2%; Pred. No. 9.4e-184;
Matches 1247; Conservative 0; Mismatches 802; Indels 93; Gaps 4;

QY 118 CAGCCTAATCACTTCGCGCATCTGTAATCGCTGCTGGCGGATTTTCCTCTCGCCT 177
Db |||||
QY 151 CACATTTTGAAGCTTCTCATCTGCAATGGCTGTGATTATCTCTTCCACTGCTCT 210
QY 178 CTTCCGCTCATTCGCTTTCGAGTCGATATTCATGAGTTTGTATCGGTTTCAACTACCG 237
Db 211 GTTCTCTCTCTGAGATTTGAAAGTGTATCCATGAGTTTGTATCCGTTTAAATTATCG 270
QY 238 GCGCACCGCTTACATGCGGAGATGTTGTCACACTTCTCAACTGTTTCCACGAGCG 297
Db 271 GACTACCAAGTTCCTGGCTGAGGAGGGTTTATATATTCATTAATCTGATGACCG 330
QY 298 CGATGATTCGCTCGGAGGATTTGGCGGTGAGCTGTATCCCGGCTGTATGATAC 357
Db 331 AGCTGGTACCTTTGGGACGAATCATTTGGAGGAACAATTTACCGAGTTTAAATGATCAG 390
QY 358 GTCCCGGCGAATCCATTTGGCTGTGACGCTACTCAACATACCGGTCCATATTCGTGACAT 417
Db 391 CTCCTGCTCAATCTACCATGTACTCTCATTTTTCACATCACCATCGACATCGGATGT 450
QY 418 CTGCTGTCTTGGCGGAGTCTTTCAGTGGCTGACCTCCATCTCCATCTCACTCTGCTGAC 477
Db 451 CTGTGTGTCTGCGCCCT 510

QY 478 CAAGAGCTGTGGTCCGGGCGCGCCCTCTTTCGCGCGCAGCTTTCATGCCCATCGTGC 537
 DB 511 CAAAGAGCTCAAGGATGACGGGCTGGCTTCTTGTCTGCCATGATTTGCTGATTTCC 570
 QY 538 TGGCTACATCAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 597
 DB 571 TGGATATATCTCCCGATCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 630
 QY 598 COTGAGTTCACCTACTTCTCTGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 657
 DB 631 CATGCTACTCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 690
 QY 658 GCGCGCAGCGCTTGTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 717
 DB 691 AGCTAAGTGTGCGCTTGTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 750
 QY 718 CATCAACCTGATACCCCTGACAGCTTCTCTGCTACTGCTCATATGCGGAGGTACTCGCGCG 777
 DB 751 GATCAACTTAATCTCTCTCCACGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 810
 QY 778 TCTGCTGACAGCTACAGCAGCTTCTACTACTACTACTACTACTACTACTACTACTACTACTACT 837
 DB 811 GATCTATGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
 QY 838 CTTGCTGGGATTCGAACCGATACGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
 DB 871 CTTTGTGGTTCAGGCTGCTCTTCTATCAGAGCAGATGGCAGGGTTTGGGCTCTTGG 930
 QY 898 GCTCCTATAGCGGTGGCCACTTTCGCCATTTGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 957
 DB 931 TCTCTGCCAGATCCATGCTTGTGGATTAACCTGCGCAGCAAGTTGAATCCCAACAATT 990
 QY 958 CCGAAGCTGTTCATCTGCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
 DB 991 TGAAGTCTTTTCCGGAGGTCATCTCTCTGCTAGGCTTGTCTCTTCTCACCGTGGAGC 1050
 QY 1018 GGTGCTCAAGTGTGGGCTTGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
 DB 1051 TCTCTCTAGTGTGACAGGAAATATCTCTCTGGAAGGGGCTTCTCTCTCTCTCTCTCTCTCT 1110
 QY 1078 TACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1137
 DB 1111 TCCCTCTTATGCTAAGAACATCCCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
 QY 1138 CACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
 DB 1171 AACCTGGCTCTACTACTATTTTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1230
 QY 1198 GTGCTACTGCTCAAGCAGATCAACGAGCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1257
 DB 1231 CTATTAAGTCTTGAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1290
 QY 1258 TCGGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
 DB 1291 CAGCATGCTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350
 QY 1318 CATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1377
 DB 1351 CATTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
 QY 1378 TAAGCGAATGGGCACAGCCATAAGCGGACCCAGGAGTGGATGAAGCTGAGGATTCAT 1437
 DB 1385 -----TCCAC 1389
 QY 1438 TGAGAAGAAGCGCTGTACGAAAGGCTGGCAAGCTGAAGCATCTGCTACTAAGCATGATGC 1497
 DB 1390 ATACATGAAGAATCTGGACATAAGTCCGCCAGACAGAGAGAGCAAGCAACAGGATTC 1449
 QY 1498 CCAGCAGGATCTGCGCTCAGCTCCACCTGAGAGATGTTGTTATTTTGGCGCTTCTAAT 1557
 DB 1450 CACCTACCTTAAAGAT-----GAAGTGGCAAGTGGGATGATGATGCTGCTGCTGCTT 1503

QY 1558 GCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
 DB 1504 CTTTCTCATCACCCTACACCTTTTCACTTCAACCTGGGTGACCACTGAGGCTTCTTCTCC 1563
 QY 1618 CTTCAATGCTTGGCTTTCACAAAGCTCAAGATGGATCCCGCAACATTTTACAGCATTT 1677
 DB 1564 GTCATTTGCTACTATCTGCCCC---GTGGTGGGATGGCAGTAGGATCATATTTGATGACT 1620
 QY 1678 CAGAGAGGCTTACTACTGCTTTCGAGAACACTGCGCATGATGCTGCTGCTGCTGCTGCTG 1737
 DB 1621 CCAGAAAGCATATATTGCTTTCGTATATCTCCAGAGGATGCAAGGCTCATGCTGCTG 1680
 QY 1738 GTGGGATACGATACAGTAGCGGAAATGGCAACAGACGAGCTAGTGGATATATA 1797
 DB 1681 GTGGGATATGGCTATCAGATTACAGCTATGGCAACCGAAACAATTTTGTGGCAATAA 1740
 QY 1798 TAGCTGGAACAATAGTCATAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
 DB 1741 CACATGGAATAATCCCATATTTCTCGAGTAGGCGAGGCAATGGCTCCACAGAGGAAAA 1800
 QY 1858 GTCTACGAAATATGACATCTCTTTCAGCTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
 DB 1801 AGCTATGAGATCATGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 QY 1918 GATCGGCTATTTCTGCGCATGATATCAACAAGTTCCTGTGGATGCTGCTGCTGCTGCTGCT 1977
 DB 1861 CACTGGGTATTTCTCTGATGATATCAACAAGTTTCTTGGATGCTGCTGCTGCTGCTGCTGCT 1920
 QY 1978 ---AGAGCATCCAGGACATTAAGGAAGGATGCTTACCGCGGCTGCTGCTGCTGCTGCTGCT 2034
 DB 1921 CACAGATACAGGCAACAATATCAAGGGAATGCTTATTAATCTCAACTGGGAGTTCCG 1980
 QY 2035 GGTAGATGCGGAGGCTTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2094
 DB 1981 TGTGGACCGTGAAGTTCCTCCAGTCTGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 QY 2095 CAGATTCGGGAAATGCAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2154
 DB 2041 TCGCTTTGACAGGTTTACAGAAAGCCAGGCTTCTCCAGGCTTTCAGCGTTGCCGAAA 2100
 QY 2155 CCGCGCTATTGGAATAAGGATTCGATCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2214
 DB 2101 TGTGAGATTGGAAATAAAGACTTTGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
 QY 2215 ACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2256
 DB 2161 ACATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2202

RESULT 10
 US-10-417-375-92
 ; Sequence 92, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: PastSeq for Windows Version 4.0
 ; SEQ ID NO 92
 ; LENGTH: 3093
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; US-10-417-375-92
 Query Match 26.1%; Score 631.2; DB 18; Length 3093;
 Best Local Similarity 58.2%; Pred. No. 3,3e-183;
 Matches 1246; Conservative 0; Mismatches 803; Indels 93; Gaps 4;
 QY 118 CAGCCTAAATCAGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177

; Sequence 9, Application US/10028384
; Publication No. US2003014285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_008408
; DATABASE ENTRY DATE: 2000-11-01
; RELEVANT RESIDUES: (1)..(3094)
US-10-028-384-9

Query Match 26.1%; Score 631.2; DB 15; Length 3094;
Best Local Similarity 58.2%; Pred. No. 3.3e-183;
Matches 1246; Conservative 0; Mismatches 803; Indels 93; Gaps 4;
QY 118 CAGCTTAATCACCTTCGCCATCTCTGTAATCGCTGGCTGGCGGGAATTTCTCTCGCT 177
DB 156 CACACTTCTAAAGCTTCTCATCTCTGTCGATGGCTGTGTATCTCTTTTCTACTCGTCT 215
QY 178 CTTCGCGCTCATCGCTTTCAGTCGATATCCATGAGTTTGTATCGCTGGTTCACATCCG 237
DB 216 TTTTGTCTGTGAGATTTGAAAGTGTCAATCCATGAGTTTGTATCGCTGTATTAATATCG 275
QY 238 GGCACCGCCTACATGTTGTCAGAAATGTTGGTACAACTCTTCACTCACTGTTTCGAGCG 297
DB 276 GACTACCGGTTTCTGCTGAGGAGGGGTTTATAAATTCCTAACTGTTTGTATGACCG 335
QY 298 CGCATGGTATCCGCTCGCGGAGGATTTGGGCGGTACCGTCTATCCCGGCTCATGATTAC 357
DB 336 GGCTTGTATCCCTTTGGGCGGATTCATTTGAGAGAACTTTTACCAGGTTTAAATGATCAC 395
QY 358 GTCCGCGGGAATTCATGTTGCTGTGACGATCTCAATACACCGGTTCCATTTCTGAGCAT 417
DB 396 TTCTGCTGCAATCACCATGTACTCCATTTCTTCCATATCACTAATGACATTCGGAATGT 455
QY 418 CTGCGTGTCTCGCGCGCATCTTTCAGTGGCTGACCTCCATCTCCACCTACCTCTGAC 477
DB 456 CTGTGTTTCTGCGCCCACTTTTCTCTCTTTTCAACCACTGTTACGTACACCTTAC 515
QY 478 CAAGAGCTGTGTCGCGGCGCGCTCTTTCGCGCGCAGCTTCATCGCCATCTGTC 537
DB 516 CAAGAGCTCAAGGATCAGGAGCTGGGCTTCTTGTGTCGATGATGTTGCTGTATTC 575
QY 538 TGGCTACATCAGTAGGTGGTGGCTGATCGTACGATTAACGAGGCAATTCATATTCG 597
DB 576 TGGGTATTTCTGATCTGTAGCTGGCTCTATGATATGAGGAAATTTGCTATCTTTG 635
QY 598 CTTGAGTTCACCTTACTTCTGTGGTGGCTCAGTGAAGACATGAGATTCGTTCTGTC 657
DB 636 CATGCTGCTTACTTACTACATGTTGATCAAGGAGTGAAGACTGGTTTCCATCTATTGGGC 695
QY 658 GCGCGCAGCGCTTTGTCTTACTTCTACATGTTGTCCGCTCGGCTGGGTTGCTGTTCAT 717
DB 696 TGCCAAAGTGGCCCTGCTTATTTCTACATGTTCTTCTATGAGGAGCTATGTTGCT 755
QY 718 CATCAACCTGATACCCCTGACGTTCTGCTGCTCATTTATGAGGAGCTATCCGCCG 777
DB 756 GATCAACTTGTATCTCTACATGTTGCTGCTGCTAAATGCTGACAGGCGGTTTCTCACCG 815
QY 778 TCTGCTGACGAGCTACAGACCTTCTACATCTCTGGGACTGCTGTCTCTCCATCGATCCC 837
DB 816 GATCTACGCTACTGTACTGTTTACTGCTGGGCGACCATTTCTTCTATCGATTTTC 875

QY 838 CTTCTGGGATTCACACCGATACGACCCAGTGAACACATGCTCGCTGGAGTGTGTGT 897
DB 876 CTTTGTGTTTCCAGCCGCTCTTTCATCAGAACACATGAGCAGCTTTTGAGTGTGTGG 935
QY 898 GTCTCTTATGCGCGTGGCCACCTTTCGCGCATTTTCAGTCCGCTGCTGCGGCAACGAGTT 957
DB 936 TCTCTGTAGATCCATGCTTCTGTAGATTTACCTCGCAGCAAGTTGAATCCACAGCAAT 995
QY 958 CCGAAGCTGTTCATCGTCGCGGATTCCTGCTGGGCTTGGGCTCTTTGTGCCGCTGT 1017
DB 996 CGAAGTCTTTTCCGAGTGTATCTCCCTGGTGGCTTTGCTCTCTCACTGTGGAGC 1055
QY 1018 GTGTCTCACCATGCTGGGCGTGTGGCCCGTGGAGTGGAGCTTCTACTCGCTGGGA 1077
DB 1056 TCTCTCTCATGCTAACAGGAAATTTCTCCCTGGACAGGGGCTTTCTACTCTCTGTGGA 1115
QY 1078 TACTGCTACGCAAGATCCACATTCCTCATCTTCATTCGCTGTCGAGCATCAGCCAC 1137
DB 1116 TCCCTCTTATGTAAGAATAACATTCCTCATTTTTCATCTCTTCTGAGCACAGCCAC 1175
QY 1138 CACTTGTCTCTGCTTCTTCTTGTATCTGCATCTCTGCTGCGCTTCCAGTGGAGT 1197
DB 1176 AACCTGGCTCTTCTACTATTTTGTATCTACAGTCTCTTGTCTCTCATTTTCCAGTTGSCCT 1235
QY 1198 GTGCTACTCATCAGCAGATCAACGACAGCGGCTTTCTGCTGCTGCTGCTGCTGCTGCT 1257
DB 1236 CTATTACTGCTTTAGCACTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295
QY 1258 TCGGCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
DB 1296 CAGCATGATCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
QY 1318 CATGCTGCGGAGTGGCTTTTTCGAGCTGTTTGGATGTTTCTGCTGCTGCTGCTGCTGCTGCT 1377
DB 1356 CATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1406
QY 1378 TAAGCAATGGGACACAGCATAAAGCGCACAGCCAGCCAGCTGAGATGAGATTCAT 1437
DB 1407 ----- 1406
QY 1438 TGGAAGAAGACCTGTAGCAAGCTGGCAAGCTGAGCATCTGCTACTAAGCATGATGC 1497
DB 1407 TCTGGACATAGTCTCCCGACAGCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1466
QY 1498 CCAGCAGGATGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1557
DB 1467 TAAGATGAGG-----TGGCGAGTGGGATGATGATGCTGCTGCTGCTGCTGCTGCTGCT 1508
QY 1558 GCTGTTGATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1617
DB 1509 TTTTCTCATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1568
QY 1618 CTCATTTGCTTGGGCTTTCCCAACAGTCAAGATGAGTCCCGCAACATTTTAGCAGATTT 1677
DB 1569 CTCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1625
QY 1678 CAGAGGCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1737
DB 1626 CGAGAGGCTGATTTTGGCTCCGTCACAACTACTTCCAGAGGATGCAAAAGTCTGCTGCTGCT 1685
QY 1738 GTGGATTAACGATACAGATAGCGGGAATGGCAACAGAACAGCAGCTAGTGGATAATAA 1797
DB 1686 GTGGATTAACGATACAGATAGCGGGAATGGCAACAGAACAGCAGCTAGTGGATAATAA 1745
QY 1798 TAGTGGACATAGTACATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1857
DB 1746 CATGGAATATACCATATTTCTGAGTAGGAGGAGCAATGCAATCCACAGAGAA 1805
QY 1858 GTCTTACGAAATATGACATCTCTGAGCTGAGCTAGCTTGTGCTGCTGCTGCTGCTGCTGCT 1917
DB 1806 AGCTATGAATCATGAGGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1865

QY 1918 GATCGCTATTCTGGGATGATATCAACAGTTCTCTGGATGGTCCGAAATTCGTGAGG 1977
 Db 1866 TACTGGGTATTTCTCGGATGATATCAACAGTTCTCTGGATGGTCCGAAATTCGTGAGG 1925
 QY 1978 ---AGAGCATCCCAAGGACATTAAGGAAAGGATTAATTAACGACCGCGGTGAATTCAG 2034
 Db 1926 CACAGAGACAGGAAGACACATTAAGGAAATGACTACTATCTCTACTGGGAATTCG 1985
 QY 2035 GGTAGATCCGGAAGGTCTCCGCGCTCTCTCACTGCTTATGTACAAATTAAGCTACTA 2094
 Db 1986 TGTGTGATCTGAGGGTCTCTCCGCTCTCTCACTGCTTATGTACAAATTAAGCTACTA 2045
 QY 2095 CAGATTCCGGGAATTAAGGTGACTTACAGAGGTCCATCTGGATATGATCGCACACGTAA 2154
 Db 2046 CGCTTTGGGAGGTCTACACAGAACCAAGCGTCCACCGCTTTGACCGTGTTCGAA 2105
 QY 2155 CGCGTCAATGGGAATAGGATCTCGCTCTCACTGCTTATGTACAAATTAAGCTACTA 2214
 Db 2106 TGTGAGATGGAATTAAGGTGACTTGAAGCTTGAAGGTGATGCTCGGAGGATATACACAGA 2165
 QY 2215 AACTGGCTTGTTCGATCTATAGGTGAAGAGCGGCATGA 2256
 Db 2166 AACTGGCTTGTTCGATCTATAGGTGAAGAGCGGCATGA 2207

RESULT 12

US-10-417-375-99
 ; Sequence 99, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 99
 ; LENGTH: 5404
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-417-375-99

Query Match 26.0%; Score 628.6; DB 18; Length 5404;
 Best Local Similarity 58.3%; Pred. No. 2,7e-182;
 Matches 1238; Conservative 0; Mismatches 794; Indels 93; Gaps 4;
 QY 118 CAGCCTAATCAGCTTCGCCATCTCTGCTAATCGCTGATCGCTGGCGGATTTCTCTCGCT 177
 Db 180 CACACTTTTGAAGCTTCTCAITCTGTCAATGGCTGCTGTATATATCTCTCCACTCGCT 239
 QY 178 CTTCGCGCTCATCCGTTTCGAGTCGATTATCCATGAGTTTGCCTGGTTCGCTGATACCG 237
 Db 240 GTTTGCTGCTCAGATTGGAAGTGTATCCATGAGTTTGCCTGATATATATATATATCG 299
 QY 238 GGCACCGCTATAGTGTGAGAAATGGTGTGATCAACTTCTCACTGCTGCTGAGAGCG 297
 Db 300 GACTACAGGTTCTCGCTGAGAGGGGTTTATTAATTCATTAATTCATTAATTCATTAAT 359
 QY 298 CGCATGGTATCCGCTCGGAGGATTTGGGGGATACCGCTATCCCGGCTGATGATTAC 357
 Db 360 AGCTGGTACCTTTGGGAGCAATCATTTGGAGCAATTTACCGAGTTTAAATGATCAC 419
 QY 358 GTCCGCGGAATTCATGCTGCTGAGTACTCAACATACCGGTCCATATTCGTGATCAT 417
 Db 420 CTCTGCTGAATCTACCATGTACTCATTTTTCACATCATCATCATCATCATCATCATCAT 479
 QY 418 CTGCGGTTCCTGGGCGGATCTTCAGTGGCTGAGCTTCCATCTCCACTTACCTGCTGAC 477
 Db 480 CTGTGTTCCTGGGCGGCT 539
 QY 478 CAAAGAGCTGTGTGTCGGGCGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537

Db 540 CAAAGAGCTCAAGGATGAGGGGCTGGGCTTCTTGTCTGCCAATGATTCGTAGTTCC 599
 QY 538 TGGCTACATCAGTAGTGGTGGATCGTACGATAAAGAGGCAATTCGCCATATTCGC 597
 Db 600 TGAATATATCTCCGATCTGTGGCTCTCTATGATAAAGAGGATTCGCCATCTTTG 659
 QY 598 CTTGAGTTACCTACTTCTCTGCTGGTGGCTCAGTGAAGACTGGATCCGCTGTTCTGGTC 657
 Db 660 CATGCTACTACCTACTACATGTTGATCAAGGAGTAAAGACTGTTCCATCTGTGGGC 719
 QY 658 GSCCGACGCGCTTGTCTCTACTTCTACATGGTCTCCGCTGGGCTGGCTGCTGCTCAT 717
 Db 720 AGCTAAGTGTGGCTTGTCTTATTTCTACATGGTCTCTCATGGAGGTTATGTGTTCT 779
 QY 718 CATGAACCTGATACCTGACGCTTCTGTAAGTCTCTATGAGGAGTACTCGCGCG 777
 Db 780 GATCAACTTAATTCCTTCCACGCTCTGCTGATGCTCACAGGCGCTTCTCTCACCG 839
 QY 778 TCTGCTGACCAAGCTACAGACCTTCTACATCTCTGGGACTGCTGTTCTCTCATGAGATCC 837
 Db 840 GATCTATGTGGCTTCTGTAAGTCTCTGCTGGGCACTATCTTCTATGAGATCTC 899
 QY 838 CTTGCTGGGATTCGAACCGATACGACAGTGAACACATGGCTGGCTGGGAGTGTGT 897
 Db 900 CTTTGTGGGTTTCCAGCCTGCTCTTTCATGAGACATGCGAGCCTTGGGGTCTTTGG 959
 QY 898 GCTCTTATGGCGGTGGCCACCTTTCGCGCAATTTGCACTGCTGCTGCGCAACGAGTT 957
 Db 960 TCTCTGCCAGATCCATGCTTGTGGATTACCTGCGCAGCAAGTTGAATCACACAAT 1019
 QY 958 CCGGAGCTGTTTCATGCTGGGAGTTCCTGCTGGGCGTGGGCTCTTTTGGCGCTCGT 1017
 Db 1020 TGAAGTCTTTCGAGCGTCTCTCTGTAAGGCTTGTCTCTCTCTCTCTCTCTCTCTCT 1079
 QY 1018 GGTGCTCACCATGCTGGCGCTTGTGGCGCGTGGAGTGAACGCTTCTCTCTCTCTCT 1077
 Db 1080 TCTCTCATGCTGACAGGAAAAATATCTCTGCGAGCGGCGTCTTCTCTCTCTCTCT 1139
 QY 1078 TACTGGCTACGCCAGATCCACATTCGCATTCGCTGCGAGCATCGCTGCGAGCATCAGCC 1137
 Db 1140 TCCCTCTTATGCTAAGAACAAACATCCCATCTCTCTCTCTCTCTCTCTCTCTCTCT 1199
 QY 1138 CACTTGGTTCGTTCTCTTCTTGTGATCTGCACATCTCTGCTGGCGCTTCCAGTGGAGT 1197
 Db 1200 AACCTGGTCTCTACTATATTTTGGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 1259
 QY 1198 GTGGTACTGCTCAAGCAGATCAAGCAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCT 1257
 Db 1260 CTATTAAGTCTTACCAACCTGCTGATGCGCGGATTTTATCATGATGATGATGATGAC 1319
 QY 1258 TGGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
 Db 1320 CAGCATGCTCTTTCAGCTGTAATGGTGGCTCTAATGCTAGTGTGGCAGCTGTTATGTG 1379
 QY 1318 CATGCTGGCGGAGTGGCGCTTCTGGGACTGTTGGATGCTGCTCTGCAAGAGGATTCGTC 1377
 Db 1380 CATCTCTCTGCGATTTGAGTCTCCAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1413
 QY 1378 TAAGGGAATGGGCAAGCAGCATTAAGCGCAGCCACCGAAGTGAATGAAGCTGAGGATTC 1437
 Db 1414 -----TCCAC 1418
 QY 1438 TGGAAGAGAGCTGTAGCAGCAAGGCTGCGCAAGCTGAAGCATGCTGTAAGCATGATGC 1497
 Db 1419 ATACATGAAGAAATCTGGAATAAGTCTGCAAGCAAGAGAGCAAGAGCAACAGGATTC 1478
 QY 1498 CCAGCAGGATATCTGGCGTCACTCCAACTGAAGAGTATTTGTTATTTTGGCGCTTCTAAT 1557
 Db 1479 CACCTACCTATT-----AAGATGAAGTGGCAAGTGGGATGATCTGCTGCTGCTGCT 1532
 QY 1558 GCTGTGTGATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1517

Db 1533 CTTTCTCATCACCTACCTTTTCATTCAACTGGGTGACAGTGAGGCTACTCTTCTCC 1592
Qy 1618 CTCGATTTCTTGGCTTTCACACAGTCAAGATGAGATCCCGCAATTTTACAGATTT 1677
Db 1593 GTCCATTTCTATCTGCCCC---GTGGTGGGGATGGCAGTAGGATCATATTTGATGACTT 1649
Qy 1678 CAGAGAGCTTACTACTGCTTTCGACAGAACACTTGCAGATGATGCTGGCTTATGCTTGG 1737
Db 1650 CCGAGAAGCATATTTATGGCTTCGTATATCTCCAGAGGATGCGAAGTCAATGCTCTG 1709
Qy 1738 GTGGGATTCAGGATACCGATAGCGGAATGGCAACAGAACGACCTAGTGGAATTA 1797
Db 1710 GTGGGATTTATGGCTATCATGATTAAGCTATGGCAACCGAACAATTTTATGAGCAATAA 1769
Qy 1798 TACGTGGAACAATAGTACATAGAGCTTGGTGGCAAGGCAATGCTTCAACCGAGGAGAA 1857
Db 1770 CACATGGAATATACCCCATATTTCTCGATAGGCGAGGCAATGGCGTCCACAGAGGAAA 1829
Qy 1858 GTCTACGAAATATGACATCTTTGAGCTGGAGTACTTTTGGTGATCTTTTGGCGGTGT 1917
Db 1830 AGCTTATGAGATCATGAGGAGCTCGATGTGAGTATGCTGCTCAATTTTGGAGGCT 1889
Qy 1918 GATCGCTTATCTGGCATGATATCAACAAAGTTCTTGGATGGTCCGAAATGCTGAGGG 1977
Db 1890 CACTGGGTATCTCTGATGATATCAACAAAGTTTCTTGGATGGTCCGGAATGGAGGAG 1949
Qy 1978 ---AGAGCATCCGAGGACATTAAGGAAGGATTTTACGACGGGCTGATTCAG 2034
Db 1950 CACAGATACGCGCAACATATCAAGGAGATGACTATTATACCTCAACTGGGAGTTCCG 2009
Qy 2035 GGTAGATCCGAAGGTGCTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTACTA 2094
Db 2010 TGTGACCGGTGAAGTCTCCAGTGTCTCACTGCTCATGTACAAGATGTTACTA 2069
Qy 2095 CAGATCCGGGAATGAAGTTGGACTACAGAGGTCCATCTGGATATGATCGCACAGTAA 2154
Db 2070 TGCCTTTGGACAGTTTACAGAAAGCCAGCGCTCCAGGCTTTGACCGTGTCCGAAA 2129
Qy 2155 CGCCCTCATTTGGGAATAGGACTTTCGATCTGACCTACCTGGAGAGGCTTACACACAGA 2214
Db 2130 TGCTGAGATTTGGGAATAAGACTTTGAGCTTGATGCTCTGGAGAGCATATACCACAGA 2189
Qy 2215 ACACGTGGCTTTTGCATCTATAGG 2239
Db 2190 ACATTTGGCTGTGACGATATACAG 2214

RESULT 13

US-10-417-375-97

; Sequence 97, Application US/10417375

; Publication No. US20040219528A1

; GENERAL INFORMATION:

; APPLICANT: David W. Morris

; APPLICANT: Marc Malandro

; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer

; FILE REFERENCE: 529452001600

; CURRENT APPLICATION NUMBER: US/10/417,375

; NUMBER OF SEQ ID NOS: 176

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 97

; LENGTH: 5827

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-417-375-97

Query Match

Best Local Similarity 26.0%; Score 628.6; DB 18; Length 5827;

Matches 1238; Conservative 0; Mismatches 794; Indels 93; Gaps 4;

Qy 118 CAGCTTAATACCTTCGCGATCTGCTAATCGGCTGGCTGGCGGATTTTCTCTCGCT 177

Db 180 CACACTTTTGAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 239

Qy 178 CTTGCGGCTACCTCGTTTCGAGTCTGATTCATCAGATTTGATCGGTGGTTCAACTACCG 237
Db 240 GTTGTGCTTCCTGAGATTTTGAAGTGTTCATCATGAGTTTTCATCGTACTTTTAATATTCG 299
Qy 238 GGCACACCGCTTACATGGTGCAGAAATGGTTGGTACAACTTCCTCAACTGGTTTCGACGAGCG 297
Db 300 GACTACACAGTTCCTGCTGAGAGGGGTTTATAAATTCATAACTGGTTTTCATGACCG 359
Qy 298 CGCATGGTATCCGCTCGGAGGATTTGGGCGGTAACCTCTATCCGCGCTGATGATTAC 357
Db 360 AGCTGGTACCTTTTGGGAGCAATCATTTGGAGGAAACAATTTACCCAGGTTTAAATGATCAC 419
Qy 358 GTCCGCGGAATCCATTTGGCTGTCGACGATCTCAACATACCGGTCCCAATTTTCGTGACAT 417
Db 420 CTCTGCTGCAATCTACCATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGT 479
Qy 418 CTGCTGTTCTTGGGCGGATCTTCAGTGGCTGATCTGATCTCCATCTCCACCTACCTGCTGAC 477
Db 480 CTGTGTGTTCTTGGGCGGCT 539
Qy 478 CAAGGAGCTGTGTTCCGCGGCGCGGCTCTTTCGCGCGCCAGCTTTCATCGCCATCGTGCC 537
Db 540 CAAGGAGCTCAAGGATGACAGGGGCTGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Qy 538 TGCTACATCAGTAGTTCGCTGGCTGGATCTGATCTGATCAACAGAGGCTTGCATATTCG 597
Db 600 TGGATATATCTCCGATCTGCTGGCTGGCTCTTATGATTAATGAAGGATTTGCACTTTTG 659
Qy 598 CTGCTGAGTTCACCTACTTCTCTGGGTGCTCTGCTGAGGACTGGATCCGTTCTGCTGTC 657
Db 660 CATGCTACTCACTACTACATGAGATCAAGGAGTAAAGACTGCTTCCATCTGTTGGG 719
Qy 658 GCGCGCAGCGGCTTGTCTCTACTTCTACATGGTCTCGCTGGGCTGGCTGCTGCTGCTCAT 717
Db 720 AGTAAATGTTGCTGCTTCTTCTTCTACATGCTCTGCTCATGGGAGGTTATGTTGCTCT 779
Qy 718 CATCAACCTGATACCCCTGCTGCTCTCTGCTGCTCTTATGAGGAGTACTCGCGCG 777
Db 780 GATCAACTTAATTCCTCTCCAGCTCTCTGCTGATGCTCACAGCGCTTTCTCTCACCG 839
Qy 778 TCTGCTGACAGGTACAGACCTTCTACATCTGGGACTGCTGCTCTCATGAGAGTCCC 837
Db 840 GATCTATGGGCTACTGTACTGTTTACTGCTGGGCTACTATCTTCTATGAGATCTC 899
Qy 838 CTTGCTGGGATTCACACCGATACGACCAAGTGAACACATGCTGCTGGGAGTGTGTTGT 897
Db 900 CTTGCTGGGTTTCCAGCTGCTCTTTCATGAGACACATGCGAGCTTTTGGGCTTTTG 959
Qy 898 GCTCCTTATGGCGGTGGCCACCTTGGGCAATTTGAGTCTGCTGCTGCTGCGCAAGATT 957
Db 960 TCTCTCCAGATCCATGCTTGTGGATTAACCTGCGAGCAAGTTGAATCCACACAAAT 1019
Qy 958 CCGGAAGCTGTTTCATCGTGGGAGTGTGCTGGGCGCTTGGCGCTCTTGTGGCGGCTG 1017
Db 1020 TGAAGTCTTTTCCGAGGCTCATCTCTGCTGAGGCTTGTCTCTCTCTCACCGTGGAGC 1079
Qy 1018 GTGCTCACCATGCTGGGCTTGTGGCGGCTGGAGTGGAGCGCTTCTACTCTGCTGGGA 1077
Db 1080 TCTCCTCATGCTGACAGGAAATAATCTCCCTGGAGGGGCTTCTACTCTGCTGCTGGA 1139
Qy 1078 TACTGGCTACGCAAGATCCATTTCCCATCATTTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
Db 1140 TCCCTCTTATGCTAGAACAAACATCCCATCATTTCTGCTGCTGCTGCTGCTGCTGCTG 1199
Qy 1138 CACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1197
Db 1200 AACTGGCTCTCATACTATTTGAGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1259
Qy 1198 GTGGTACTCATCAAGCAGATCAACGAGCGGCTTTTCTGCTGCTGCTGCTGCTGCTGCTG 1257
Db 1260 CTATTAATGCTTAGCAACCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1319

QY 899 CTCCTATATGGCGTGGCCACCTTGGGCCATTGTCAGTCCGTCGTCGTCGGCGCAAGAGTTTC 958
DB 1006 CTGTGTACAGCTGATCGGATTCTCGAAGTCGTCCGACGACTCGTGGCTGGCGAAGCAATTC 1065
QY 959 CGGAAGCTGTTTCATCGTGGCGGATTGCTGGTGGCGCTTGGCTCTTTGGGCGGTCGTCG 1018
DB 1066 CAGCTCTTCTCAAGCTTTTGTGTCGGCGTATTCTGCGCTCAGTTTGTCTGCTCCCTGTC 1125
QY 1019 GTGCTCACACACTGCTGGCGGTTGTGGCCCGGTCGAGTGGACGCTTCTACTGCTGTGGAT 1078
DB 1126 ACTTTGACTTTCTCTGGATGATCGCCCTTCTGCTGGAAGATTTTATTCTCTTTGGAT 1185
QY 1079 ACTGCTACGCCCAAGATCCACATCCCATCAATTCATCCGTCGTCGAGCATCAGCCACC 1138
DB 1186 ACTGCTATGCGAAGTCCACATGCCCATATTGCTCTCGTCTCCGACACACAGCCACC 1245
QY 1139 ACTTGGTCTGTTCTTCTTTGTGATCTGCACATCTCTGGTGTGGCTTCCAGTGGAGTG 1198
DB 1246 GCTTGGCCCTCATTTCTACTTTGACTCCTCAAAATGCTTATCTTCTTTTCCCTGCGGTGTC 1305
QY 1199 TGGTACTGCATCAAGCAGATCAACGACGAGCGCTTTTCTGCTGTGTCAGCCATCAT 1258
DB 1306 TTTCTGTTTCAAGGAGTTCGCGATGACAGATCTTCAATCATATTATGCGGTCCTC 1365
QY 1259 GGGTTTACTTCTGCTGTGATGCTGGTGTGATGTTGATGACCTCTACGCGGTGGTGTG 1318
DB 1366 ATGCTCTATTGTCGGGTGTCATGTTGCACTTATGCTTGTATCATCAGCTGTGTTGCTGT 1425
QY 1319 ATGCTGGCGGAGTGGCTTTTCCGACTGTGATGTTGCTGCAAGAGATTCGCT 1378
DB 1426 GTTCTCTCGCAATGCTTCTCCAACTTCTCGAGCGTATATTGACCGCTCATCCCC 1485
QY 1379 AAGCAATGGGACACGATCAAGCGGACGACCGAAGTGTGATGATGAGCTGAGATTCAT 1438
DB 1486 GAAACGACGAGGAGCTGGCGAGTCTCAGACGAGGTTGCTCTCAAGTCCAAGCGAAG 1545
QY 1439 GAGAAAGAGAGCGTGTACGACAGCGCTGGCAAGCTGGAAGCATGTGACTTAAGCATGATGC 1498
DB 1546 AAGATGGCGCTGCGCAACGCGCAATAAGAGCGGTCTCTTTACAGGTATTTGAGCGC 1605
QY 1499 CAGCAGGATACGGGTCAGCTCAACCTGAGAGTATGTTATTTTGGCGGTTCTAATG 1558
DB 1606 AAGTCTGCTCGGATCTTTGGTCTCGACATCGATTTGCTGTGTTTCCATTTCTCT 1665
QY 1559 CTGTTGATGATGTTCTGCTGTC---CACTGACAGTGGGTGACCAAGTCTACTCCAGT 1615
DB 1666 GTCTTCTCTTCACTTTTGTCTTCACTGACATATGTGACTTCAACAGGTATTTCTG 1725
QY 1616 CCTCATGTTGCTGGCTTCCACACAGTCAAGATGGATCCGCAACATTTTAGAGAT 1675
DB 1726 CTTTCACTGATGCTTGTATCGGAAC---CCGATGTTAGGCAAAATATCATTTGATGAT 1782
QY 1676 TTCAGAGAGGTTTACTGCTGCTTTCGCAAGAACACTGCCGATGATGCTGGCTTATGCT 1735
DB 1783 TTCAGAGAGGTTTACTGATGCTTTCGCAAGAACACACGCGGACAGAGGTCATGCTC 1842
QY 1736 TGGTGGATTAGGATACAGATACGGGATGCGCAACAGACAGCTAGTGGATAT 1795
DB 1843 TGGTGGATTAGGATACAGATACGGGATGCGGATGATGCTGATCGCCCACTTTGTTGATA 1902
QY 1796 AATAGTGGAAATATGATCAATAGCTGCTGTTGGCAAGGCAATGCTTCAACCGAGGAG 1855
DB 1903 AATAGTGGAAATATGATCAATAGCTGCTGTTGGCAAGGCAATGCTTCAACCGAGGAG 1962
QY 1856 AAGTCTAGAAATATGATCTTCTGAGTGGACTACGTTTGTGATCTTTGCGGT 1915
DB 1963 GTGCGATATCTTCTTGGGAAGCATGATGCTGATTTCTTGTGATCTTTGGGCG 2022
QY 1916 GTGATGCGCTATTTGCGGATGATATCAACAGTTCCTGTGATGTTCCGAATTCCTGAG 1975
DB 2023 TTATTTGGGCTACTCTGTGTAGGATATCAACAGTTTTGTGGATGTTAGGATCTCACA 2082

QY 1976 GGAGAGATCCCAAGGACATTAAGGAAGCGATTACTTTCCGACCGCGGTGAATTCAGG 2035
DB 2083 GGTGAATGGCTGACAGGTGAGGAGTCAACTACTTTTCTCAAGAGGGGAGTATGCT 2142
QY 2036 GTAGATCCGGAAGGTGCTCCGGCCCTGCTCAACTGCTTATGTATCAAAATTAAGTACTAC 2095
DB 2143 GTCGATGACAGGCGCACCCCTACTATGAAGAACTCTCTCATGTACAAATGCTTACTAC 2202
QY 2096 AGATTCCGGGAATTGAAGTTGAGTACAGAGGTTCATCTGATATGATCGCACAGTAAC 2155
DB 2203 CGCTTCCCGA-----GCTTTATGTTGAGACACCCGCTCAAGACAGGTTGAGGC 2253
QY 2156 GCCTCATTTGGGAATAAGGACTTCGATCTGACTCTACCTGGAGGAGGCTTACACACAGAA 2215
DB 2254 CAAATATCCCTCTAAACAGTGTACTCTTGATATCTCTTGACGAAGGTTCCATATCCGAA 2313
QY 2216 CACTGCTTGTTCGATCTATAGGTTGAAGAG 2248
DB 2314 AATTGATCGTCAGGATCTACAAAGTCAAGAAG 2346

RESULT 15
US-10-425-115-150745
; Sequence 150745, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 150745
; LENGTH: 2839
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MR14577_59000C.1
US-10-425-115-150745

Query Match 25.6%; Score 619; DB 18; Length 2839;
Best Local Similarity 58.5%; Pred. No. 1.8e-179;
Matches 1244; Conservative 0; Mismatches 820; Indels 63; Gaps 7;

QY 121 CCTAATCACTTCGCACTCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCCTCTT 180
DB 283 CTTGCTCCGGCTCTCGGCTGGCGCTCATCTACGTCTCGCTCGCCTGCTCGCCTCTT 342
QY 181 CGCGCTCATCGGTTTTCGATCGATTCATCATGATTTGATCCGTGGTTCACACTACCGGC 240
DB 343 CTCCGCTCTCGCTACGAGTCCATGATCCACAGTTTCGACCTTACTTCAACTACCGCAC 402
QY 241 CACGCGCTACATGTTGTCAGAAATGTTGTGTACAACTTCTCACTGCTTCGACGAGCGCG 300
DB 403 CACGCTCTTCTCACCAGAACGGCTTCAGCGAGTTCTGGAACCTGTTTCGACTTCGAGAG 462
QY 301 ATGGTATCCGCTCGGAGGATTTGCGGCGGTAACGCTCTATCCCGCCTGATGATGATGTC 360
DB 463 CTGGTACCGCTCGGCGGCTTGTGGGGGAGCGCTCTACCGGCTGTGATGGTGAACCGC 522
QY 361 CGCGGGAATCCATTTGCTGCTGCTGCTGCTCAACATACCGGTCCATATTCGTGACATCTG 420
DB 523 CGCGCTGCTTCCACCGCTCTCTCGCGGCTCTACTCTACCGTCCACATCTCGTGAAGTCTG 582
QY 421 CGTGTTCCTGGCGCGGATCTTCAGTGGCTGAGCTGCAATCTCCACCTACCTGCTGACCAA 480
DB 583 CGTCTCAACCGCGGCGCTTCTGCTGCTGCAACACCAACGCTGCTGCTACGCTTCGCGCG 642
QY 481 GGAGCTGTGTGTCGCGGCGGCGCGCTCTTTCGCGCCAGCTTTCATCGCCATCGTGCCTGG 540

Db 643 CGAGATATGGGACTCGGGGGGGGCTCGTCCCGCTCGCTGATCGCCATCGTCCCTGG 702
QY 541 CTACATCAGTAGTGGTGGTGGATCGTACGATAAAGAGGGCATTTGGCATATTCGCGCT 600
Db 703 CTACATCTCCGCTCGTGGGGGTCTACGACAAAGAGGGCGTGGCCATCTCGGCT 762
QY 601 GCAGTTCACTACTTCTCTGGGGTGGCTCAGTGAAGACTGGATCCGGTGTCTGGTGGC 660
Db 763 GTGTCTACCTTCTACCTCTCGTGGCGCGGTTAACTCGGTGGCTCGCTGGGGGCT 822
QY 661 CGCAGCGGCTTGTCTCTACTTCTACATGGTGGTCCCGCTGGGTGGCTACGTTGTTCACT 720
Db 823 CGCTCGGGCTTCGGTTACTTCTATATGGTGGTGGCATGGGAGGCTACGTTCACT 882
QY 721 CAACCTGATACCCCTGACGCTTCTGTAATGATGCTGATGATGAGGAGGCTACGCGCGCT 780
Db 883 CAATCTCGTCCCGCTTATGTGCTTGTGCTGCTGCTGAGGAGTACTCGCAGAGGCT 942
QY 781 GCTGACGAGTACAGCACCTTCTACATCTGGGAGTGTCTCTCCATGACAGATCCCTT 840
Db 943 CTACGTGGCTACAAATTCATGATGCTGGGATGTTGCTTGCATGACAGATCGTT 1002
QY 841 CGTGGGATTCGAACGATACGACAGGATGACATGCTGCTGGCTGGGAGTGTGCT 900
Db 1003 TGTGGATTCAGCATGTGACCTCGGAGAGCACATGCTGCTATGGAGTCTTCTTCT 1062
QY 901 CTTATGGCGGTGGGCGACCTTGGCGCATTTGACGTCCGTGCTGCGCAACGAGTCCG 960
Db 1063 GTTGCAGGTTTCTTCTTCTGACTGGGTGAATATCTGTAATGATGCCAGACTAT 1122
QY 961 GAAGCTGTTCAT---CGTGGCGGATGCTGGTGGCGGCTTGGCTGTTTGGCGGCTCGT 1017
Db 1123 CAAGTCATTCCTCGAATTAACCTCACAATGTGATTAATGCTGACCTCGCTGCTGG 1182
QY 1018 GGTGCTCACATGCTGGGGGTTGTGGCGCGGTGGAGTGAGCTCTTACTCGCTGGGA 1077
Db 1183 ACTTGTACTGATCTGGTTACATCTCCCTTGGACAGAGCGTTTACTCCCTGCTGA 1242
QY 1078 TACTGCTACGCAAGATCCACATTCGATATTCATTCGCTGGTGGAGCATAGCCAC 1137
Db 1243 TCTACTACGCAAGGACCATATACCAATCATCATCTGTTCTGAGCATAGCCAAC 1302
QY 1138 CACTTGGTCTCGTCTTCTTTCATCTGCATCTGCTGGTGGCGCTTCCAGTGGGAGT 1197
Db 1303 AGATGCTGCTCTTCTCATGTTGACTTCCATCTTCTTCTTCTTCCAGCTGGCT 1362
QY 1198 GTGTAATGATCAAGCAGATCAAGAGGCGGCTTGTGGTGGTGGTGGTGGTGGTGG 1257
Db 1363 TTATTTCTGTTTCAAGCGCTGTGAGATGCCAATATTTATTTATGATGATGCTCCAC 1422
QY 1258 TGGGTTTACTTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1317
Db 1423 AAGTATGTAATTTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1482
QY 1318 CATGCTGGCGGAGTGGGCTTTTTCGGGAGTGTGGATGTGTTCTGCAAGAGGATTCGTC 1377
Db 1483 CTTTATTT-----AGTGTATTTGCTGCACTGCTGCAATTAATAAATAA 1522
QY 1378 TAAGCGAATGGGCAGACCATTAAGCGCAGCCAGCCAGTGGATGAAGTGGAGTTCAT 1437
Db 1523 CTAAACCATTTGATCAGGACAAAGAGAGAGTCCA-----CAGACTACTTCTG 1571
QY 1438 TGAGAGAGAGCGCTGTACGACAAAGCTGGCAGCTGAAGCATCGTACTAAGCATGATGC 1497
Db 1572 GGAAGAAACAACTGGCTCAAGTCA-----GCTGCAAGAGGCTGAGT 1611
QY 1498 CCAGCAGGATACCTGGGCTCAGCTCCAACTGAAGAGTATGTTATTTTGGCGCTTCAAT 1557
Db 1612 TGATCAATCTTACCATTCACACATGCTGCTATGCTTCTTCTTGGTGGCTTTTA 1671
QY 1558 GCTGTTGATGATGTTCTGCTGTCCACTGCACTGGGTGGTGGTGGTGGTGGTGGTGG 1617

Db 1672 CTGTCTAGTAGGTATGCCATACACTGCTGGGTGACATCTGAGGCTTACTCTTCTCC 1731
QY 1618 CTCCATTTCTTGGCTTTTCCAAACAGTCAAGATGGATCCCGCAACATTTTAGACGATTT 1677
Db 1732 ATCTATAGTTTGGC---CGCAAGGGGTCAATAATGGTGAAGAGTCACTCTTGGATTA 1788
QY 1678 CAGAGAGGCTTACTACTTGGCTTTCCAGAACACATGCGGATGATGCTCGCTTATGCTTG 1737
Db 1789 TCGTGAAGGATACATTTGGCTTCGTGAGAACACGCTCCCTGATGCTAAGATTTATGCTTG 1848
QY 1738 GTGGGATTTACGATACAGATAGCGGAAATGGCAAAACAGACGAGCTAGTGGATATAA 1797
Db 1849 GTGGGACTATGATACCAAAATACAGCTATGGGTAAACAGAACTGTTATTTGATACAA 1908
QY 1798 TACGTGGAAACATAGTCAATAGCTGGTGGTGGCAAGCAATGCTTCAACCGAGGAA 1857
Db 1909 TACATGGGAATACACACATAGCTACAGTTGGAGAGCTATGTCATCTTATGAAGATGA 1968
QY 1858 GTCTTACGAAATTTAGCATCTTGTAGCTGGACTAGCTTTTGGTGTATCTTTGGCGGTGT 1917
Db 1969 AGCATATGAGATCATGCACTCATGATGTGAACATATGCTTGTGTTGGAGGTGT 2028
QY 1918 GATCGCTATTTGGCGGATGATCAACAGTTCCTGCTGGATGCTCGAATTCGTGAGGG 1977
Db 2029 TACTGGATTTCTTCAGATGATATCAACAGTTCTTATGATGGTACGTATTTGGTGGAG 2088
QY 1978 AGAGATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCGACCGCGTGAATTCAGGGT 2037
Db 2089 AGTTTTTCC---TGTAATCAAAGAGCCAGATTACCTT---GTTAATGGGAGTATCGTAT 2142
QY 2038 AGATCCGAGAGTGTCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAG 2097
Db 2143 TGACAAAGGGGGCAGCACCCAAAATGCTGAATGCTGCTAATGTACAAAGCTTTGTACTACCG 2202
QY 2098 ATTCGGGAAATGAGTTGGACTACAGAGTCCATCTGGATATCATCGCACACGTAAACGC 2157
Db 2203 ATTTGGAGACTTTACACAGAAATATGGAAAACCTCCAGGGTATGATCGTGTACGAGAGT 2262
QY 2158 CGTCAATTTGGGAATAGGACTTCGATCTGACCTACTCGAGGAGGCTTACACACAGAA 2217
Db 2263 GGAGATTTGGCAACAAAGACATAAAGCTTGAATACTTTGGAGAGGCAATTTACAACTTCAA 2322
QY 2218 CTGGCTTTGTTGGCATCTATAGGGTAA 2244
Db 2323 CTGGATAGTGGCATATACAGGTAA 2349

Search completed: December 15, 2004, 17:14:06
Job time : 1235.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2004, 00:24:44 ; Search time 7349.2 Seconds
(without alignments)
11984.275 Million cell updates/sec

Title: US-10-028-384-7
Perfect score: 2417
Sequence: 1 tctaagcaagaatgtgtcg.....ccaaaaaataaaaaa 2417

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Query Length | DB ID | Description |
|------------|-------|-------|--------------|------------|--------------------|
| 1 | 960.6 | 39.7 | 2709 | 3 AK018758 | Mus muscu |
| 2 | 945.6 | 39.1 | 4513 | 3 BC044321 | Xenopus 1 |
| 3 | 944 | 39.1 | 2669 | 3 AK012153 | Mus muscu |
| 4 | 817.2 | 33.8 | 836 | 9 CNS0172X | AL108711 Drosophil |
| 5 | 807.4 | 33.4 | 810 | 2 BF502026 | AT17657.5 |
| 6 | 711.8 | 29.4 | 715 | 7 CK657451 | LP23759.5 |
| 7 | 710.8 | 29.4 | 736 | 1 AF949890 | LD39946.5 |
| 8 | 699.4 | 28.9 | 701 | 2 BF503420 | AT19406.5 |
| 9 | 688.4 | 28.5 | 690 | 4 BG540860 | SD12024.5 |
| 10 | 681 | 28.2 | 681 | 7 CK659033 | LP16462.5 |
| 11 | 680.4 | 28.2 | 694 | 4 B1354111 | GM26338.5 |
| 12 | 677.4 | 28.0 | 679 | 7 CK657649 | LP24119.5 |
| 13 | 676.2 | 28.0 | 682 | 7 CK659064 | LP16523.5 |
| 14 | 659.8 | 27.3 | 663 | 4 BG641172 | SD12448.5 |
| 15 | 658 | 27.2 | 666 | 4 B1357074 | RE34325.5 |
| 16 | 658 | 27.2 | 668 | 4 B1227902 | RE32588.5 |
| 17 | 653.4 | 27.0 | 655 | 2 BF496296 | AT10060.5 |
| 18 | 652 | 27.0 | 652 | 4 B1171940 | RE13841.5 |
| 19 | 648.4 | 26.8 | 655 | 4 B1484774 | RE57926.5 |
| 20 | 644.4 | 26.7 | 658 | 4 B1374189 | RE61718.5 |
| 21 | 644 | 26.6 | 644 | 4 BG636414 | SD41423.5 |
| 22 | 644 | 26.6 | 654 | 4 B1374334 | RE61893.5 |
| 23 | 641 | 26.5 | 641 | 1 A1257750 | LP06212.5 |
| 24 | 635 | 26.3 | 642 | 4 B1173041 | RE15774.5 |

| | | | | | |
|----|-------|------|------|------------|-----------|
| 25 | 633.4 | 26.2 | 635 | 1 A1135629 | |
| 26 | 631.2 | 26.1 | 3802 | 3 AK030363 | Mus muscu |
| 27 | 630 | 26.0 | 646 | 4 B1170481 | RE11944.5 |
| 28 | 628.8 | 26.0 | 2656 | 3 BC028897 | Mus muscu |
| 29 | 628.4 | 26.0 | 630 | 1 A1295381 | LP08987.5 |
| 30 | 625.6 | 25.9 | 649 | 4 BG641064 | SD12296.5 |
| 31 | 623.6 | 25.8 | 649 | 4 B1162865 | RR02124.5 |
| 32 | 621.8 | 25.7 | 2730 | 3 AK081547 | Mus muscu |
| 33 | 620.8 | 25.7 | 3643 | 3 AK077877 | Mus muscu |
| 34 | 614.4 | 25.4 | 616 | 2 BF504378 | AT05729.5 |
| 35 | 612 | 25.3 | 615 | 4 B1364613 | RE49773.5 |
| 36 | 603.2 | 25.0 | 2118 | 9 AY418284 | Hom sapi |
| 37 | 600 | 24.8 | 600 | 4 B1163592 | RR03026.5 |
| 38 | 596.4 | 24.7 | 598 | 4 B1170396 | RE11825.5 |
| 39 | 585 | 24.2 | 585 | 1 A1133998 | GH11327.5 |
| 40 | 580.6 | 24.0 | 2727 | 3 AK087470 | Mus muscu |
| 41 | 576 | 23.8 | 585 | 4 B1242003 | RE39217.5 |
| 42 | 568.6 | 23.5 | 590 | 1 AA694951 | GM01838.5 |
| 43 | 568.2 | 23.5 | 2103 | 9 AY418286 | Mus muscu |
| 44 | 562.4 | 23.3 | 564 | 1 A1542327 | SD08615.5 |
| 45 | 562 | 23.3 | 564 | 4 B1635692 | SD17119.5 |

ALIGNMENTS

| | | | | |
|------------|--|-------------|---|-----------------|
| RESULT 1 | AK018758 | 2709 bp | mus musculus adult male liver cDNA, RIKEN full-length cDNA library, clone:1300006C19 product:hypothetical Oligosaccharyl transferase (Oraee) STT3 subunit containing protein, full insert sequence. | HTC 03-APR-2004 |
| LOCUS | AK018758 | | | |
| DEFINITION | AK018758.2 | GI:26384577 | HTC; CAP trapper. | |
| ACCESSION | AK018758 | | Mus musculus (house mouse) | |
| VERSION | AK018758.2 | | | |
| KEYWORDS | HTC; CAP trapper. | | | |
| SOURCE | Mus musculus | | | |
| ORGANISM | Mus musculus | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| AUTHORS | Carninci, P. and Hayashizaki, Y. | | | |
| TITLE | High-efficiency full-length cDNA cloning | | | |
| JOURNAL | Meth. Enzymol. 303, 19-44 (1999) | | | |
| MEDLINE | 99279253 | | | |
| PUBMED | 10349636 | | | |
| REFERENCE | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) | | | |
| AUTHORS | Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Iehikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, N., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) | | | |
| TITLE | RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer | | | |
| JOURNAL | Genome Res. 10 (11), 1757-1771 (2000) | | | |
| MEDLINE | 20530913 | | | |
| PUBMED | 11076861 | | | |
| AUTHORS | The RIKEN Genome Exploration Research Group Phase II Team and the PANTON Consortium. | | | |
| TITLE | Functional annotation of a full-length mouse cDNA collection | | | |
| JOURNAL | Nature 409, 685-690 (2001) | | | |
| REFERENCE | | | | |

QY 1002 TCTTTGTCGGCGTGGTCTGCTACCATGCTGGGGGTTGTGGCCCGCTGGAGTGGACGCT 1061
 Db 1157 TGTTCCTTAGTGTCTATCTGACATACACAGGTATATTGACCACTGGAGTGGCAGGT 1216
 QY 1062 TCTACTCGCTCTGGGATACCTGGTACGCAAGATCCCAATTCCTCCATCATTTGCATCCGCTGT 1121
 Db 1217 TTTATTCACTATGGGATACCTGGTATGCAAAATACACATTTCCCAATTTATGCACTGT 1276
 QY 1122 CGGAGCATCAGCCACCACTTGGTCTCGTTCTTCTTGTGATCTGCACATCCTGGTGGCG 1181
 Db 1277 CTGAACATCAGCCATCAGCATCGGTGGTCTTCTTCTTGTGATCTACATATTCTTGTATGTA 1336
 QY 1182 CTTTCCCACTGGGAGTGGTACTGTCATCAAGCAGATCAACGAGCGGGTCTTCTGGTG 1241
 Db 1337 CTTTCCCACTGGGAGTGGTACTGTCATCAAAATATCAAGATGAAGAGTATTGTGCG 1396
 QY 1242 TGTGTAGCCCATCAGTGGGTTTACTTTCGCTGTGTGTGATGGTGGTTCATGTTGACCC 1301
 Db 1397 CTCTGTATGCGATCAGTCTGTGTACTTTCGCGGAGTGTGGTGGGCTGATGCTGACTC 1456
 QY 1302 TCAGCCCGGTGGTGTGATGCTGCGGAGTGGCTTTTCGGGACTGTTGGATGCTGTCC 1361
 Db 1457 TGACCCCGGTGGTGTGATGCTGCGGATGCGCTTCTCAATGTTTTTGGGACTATT 1516
 QY 1362 TCGAAGAGGATTCTCTAAGCGAATGGGACAGCCATTAAGCGCAGCCACCGAAGTGGATG 1421
 Db 1517 TGGGGGATGA-----CATGAAAGGGAACCCCACTGTGGAG-----GACAGCAGTG 1564
 QY 1422 AAGCTGAGGATTCATTGAGAAGAGCGGTGACGACAGGCTGGGCAAGCTGAAGATC 1481
 Db 1565 ATGAGGATGACAAAGAAACCCAGGAACCTTTGATGACAGGCGGTAAAGTGGAGAAC 1624
 QY 1482 GTACTAAGCATGATGCCAGCAGGATATCTGGCGTCAAGTCCAACTCAAGAGTATTGTTA 1541
 Db 1625 ATGTGACAGAGCAAGAGAAACCTGAAGAGGCTTGGSCCCCAACATCAAAAGCAITGTGA 1684
 QY 1542 TTTTGGCGGTTCAATCTGTTGATGATGTTGCTGTGCTCACTGAGTGGGTGACCAAGA 1601
 Db 1685 CCATGCTGATGCTATGCTCTGATGATGTTGCGGTCCACTGCACTGGGTGACCAAGA 1744
 QY 1602 ATGCTACTCCAGTCCCTCCATTGCTTGTGCTTTCACACACAGTCAAGATGGATCCGGA 1661
 Db 1745 ACGCTACTCCAGTCCCAAGTGTGCTCTGCTCTCTACAA---TCATGATGGTACCAGGA 1801
 QY 1662 ACATTTAGAGATTCCAGAGGCTTACTTGTGCTTTCGAGAACACTGCCGATGATG 1721
 Db 1802 ATATATTAGATGATTTTAGAAGAGCGTACTTTTGGCTGAGACAAACACCGATGAACAG 1861
 QY 1722 CTCGCTTATGCTTGTGGGATTTACGATACAGATAGCGGAATGGCAACAGAACGA 1781
 Db 1862 CCGGGTCACTGCTGGTGGGATCTACGCTATCAGATTTGTCGATGGCCCAACAGGACCA 1921
 QY 1782 CGCTAGTGGATAATACTGGAACAATAGTCCACATAGCGCTGTGGTGGCAAGGCAATGT 1841
 Db 1922 CTCTGTGGGATAACAACCTCGAACAACAGCCACATCGCACTGGTGGGAAAGCTATGT 1981
 QY 1842 CTTCAACCGAGGAGAGTCTAGCAATATGATCATCTCTTGTGAGTGGACTAGTGGTGG 1901
 Db 1982 CTTCCATGAACCGCGCTTATAAAATCATGAGTCCCTGTGATGTCGATATTGTGTGG 2041
 QY 1902 TGATCTTTGGCGGTGATCGGCTATTCTGGCGATGATATCAACAGTTCCTGTGGATGG 1961
 Db 2042 TTATTTTCGAGGAGTATGGCTATTTCGGGAGCGATATCAACAGTTCCTGTGGATGG 2101
 QY 1962 TCCGATTTGTAGGAGAGCATCCCAAGGACATTAAGGAAGCGATTAATTTCACGACC 2021
 Db 2102 TCAGGATAGCTGAAGGGGAGCATCCCAAGACATCCGGAAGGTGACTATTTCAACCCAGC 2161
 QY 2022 CGCGTGAATTCAGGTAGATCCGGAAGTGTCTCCGSCCTGTCTCACTCCCTTATGTACA 2081
 Db 2162 AGGGAGGATTCAGGTAGACAAAGTGGTCTCTCTACTCTGTTAACTGCCCTATGTATA 2221
 QY 2082 AATTAAGCTACTACAGATTCCGGGAATTAAGTTGGATACAGAGGTCCATCTGGATATG 2141

Db 2222 AATGTCTACTACAGATTGGAGAAATCGAGTAGATTTTCGCACTCCCCAGGCTTTG 2281
 QY 2142 ATCGCACAGTAACCGCGTCAATTGGGATAAAGGACTTCGATGACCTACCTGGAGGAG 2201
 Db 2282 ACGAAACAGTATGCTGAGATTGGAATTAAGACATTAATTCAGCATTTGGAGGAAG 2341
 QY 2202 CTTACACCAAGAACACATCGCTTGTTCGATCTATAGGTGAAGAGCGCATGAGTTCA 2261
 Db 2342 CTTTATCATCAGAGCAGCTGGCTGTGAGGATATATAAAGTGAAGCAACCTGACACAGG 2401
 QY 2262 ATAGACCATCAGTGAAGCAAGGAGAGAACCATTCCTCCAGCAACTTCATTTCCAGAA 2321
 Db 2402 AGACCTAGTGCACAAACCTCGAGTCAACCAATCGTCCCAACAGAAATTTGTCAA 2461
 QY 2322 AGAAGCTTAAGCGTGCAGAG---GGCTACATACAGAAACCGCGGTGTGTTAAGGAA 2378
 Db 2462 AGAAGACTACTAAAGGAGCGTGTAGTTAAATAAAGTACTGTGTTAAGAAAGCA 2521
 QY 2379 AACGACCTTGAATAAACCACCAAAAA 2405
 Db 2522 AGAAGACCTCTAAGAAGACTGTTTAAA 2548

RESULT 2

BC044321
 LOCUS 4513 bp mRNA linear HTC 23-JAN-2003
 DEFINITION Xenopus laevis, clone IMAGE:4684177, mRNA.
 ACCESSION BC044321
 VERSION BC044321.1 GI:27882209
 KEYWORDS HTC.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopus laevis; Xenopus; Xenopus.
 REFERENCE
 1 (bases 1 to 4513)
 AUTHORS Klein,S. and Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene
 Collection (XGC), National Institute of Child Health and Human
 Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
 20892-7510, USA
 REMARK
 NIH-MGC Project
 Contact: XGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. Igor Dawid
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 94 Row: i Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, similarity but not identity to protein
 This clone has the following problem: frame shifted.

FEATURES

source

1..4513
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:4684177"
 /tissue_type="Embryo, stage 31/32, Xenopus"
 /clone_lib="NICHD_XGC_Emb4"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

| | | | | | | | |
|-----------------------|--------|---|---------------|------------|------|---|-------|
| Query Match | 39.1%; | Score | 945.6; | DB | 3; | Length | 4513; |
| Best Local Similarity | 64.4%; | Pred. | No. 8.5e-246; | Indels | 26; | Gaps | 4; |
| Matches | 1491; | Conservative | 0; | Mismatches | 799; | | |
| QY | 102 | AGTGGCTGGCTACAGCAGCTTAATCAGCTTCGGCATCTCTGATTAATCGCTGGCTGGCGG | 161 | QY | 1122 | CGAGCATCAGCCACCACTTGGTTCTCGTTCTCTTTTGATCTGCAATCTCGTGGCG | 1181 |
| DB | 446 | AGCCGGGGCTGGCAGTCTCTGCTCTCTTCACTATTCTATTCTCGCTGGCTGGCGG | 505 | DB | 1466 | CTGAACATCAACCAACATGGGTGCTCTCTTTTGATCTGATATATAGTATGTA | 1525 |
| QY | 162 | GAATTTCTCTGCTCTTTCGGCGTATCCGTTTCGATGCTGATTAATCCATGATTTGATC | 221 | QY | 1182 | CTTCCAGTGGAGTGGTACTGCAATCAAGCAGATCAACAGAGAGCGCTTTTCGTGG | 1241 |
| DB | 506 | GCTTCAGCTCAAGGCTCTTCGCTGTATCCGCTTCGAGAGCATCATCCAGATTCGACC | 565 | DB | 1526 | CATTTCTGTCAGGACTTGGTTCTGCATAAAATATCAACONTGAAGAGTTTGTGG | 1585 |
| QY | 222 | CGTGGTCAACTACCGGCCACCGCTACATGCTGCAGATGGTTGGTACAACTTCCTCA | 281 | QY | 1242 | TGCTGACCCCATCAGTGGCTTTACTTCGCTGGTGTGATGGTGGTTGATGTGACCC | 1301 |
| DB | 566 | CTGGTTTAACATACAGATCTACATCAGCTAGCATCCATCGCTTTTATGAATTTAA | 625 | DB | 1586 | CTCTGATGCAATAGTGGCTTTACTTTGCTGGGCTAATGGTTGCTTGTGCTGACCT | 1645 |
| QY | 282 | ACTGTTTCAGAGCGCGCATGATATCCGCTCGGAGGATTTGGGGCGGTACCGTCTATC | 341 | QY | 1302 | TCAGCCGGTGGTGTGCTGCTGGCGGAGTGGCTTTTCGGGACTGTTGGATGTGTTCC | 1361 |
| DB | 626 | ATTGGTTTGAACCGGCATGCTATCCACTGGGAGAAATAGTAGGTGGAACAGTATACC | 685 | DB | 1646 | TGACTCTGTTGGCTGCTGGGCAATGGCTTCTCCAATGTTTGTGAGCAATATC | 1705 |
| QY | 342 | CGGCTGATGATAGTACGTCGGCGGAATCCATTGGCTGTCAGCTACTCAACATACCG | 401 | QY | 1362 | TGCAAGAGGATTCGTTAAGCGAATGGGCACAGCCATTAAGCGCAGCCACCGAAGTGGATG | 1421 |
| DB | 686 | CTGGATTAATGTTACAGCTGGGCTCATCCATTGGAATTTAAAGATTAACATTAACG | 745 | DB | 1706 | TGGGTGATGATATGAAGCGGGAACCCCGCAGTGGGAAGACAGCAGTGCAGAGAGCGAA | 1765 |
| QY | 402 | TCCATATTCGTGACATCTCGCTGTTCTCGGCGCGGATCTTCAGTGGCGTGCCTCCATCT | 461 | QY | 1422 | AAGCTGAGGATTCCTATGAGAAAGACGCTGTAGCAAGGCTGGCAAGCTGAGAGATC | 1481 |
| DB | 746 | TTCATATACAGATGATGTGTGTTCTTAGCACCAAGTATTTAGTGGCTTACATCAATCT | 805 | DB | 1766 | ATTCTG-----GTACACTCTATGATAAGGCTGGCAAGTGAAGAGC | 1807 |
| QY | 462 | CCACTACTCTGACAGAGCTGTGTGCGGGCGCGGCTCTTCGCGGCGCTTCGCGGCGCT | 521 | QY | 1482 | GTACTAAGCATGATGCCAGCAGGATCTGGGCTCAGCTCCCACTGGAAGATATTGTTA | 1541 |
| DB | 806 | CCACTTCTCTGCTCACTCAGAACTGTGGAACAGGAGCGGGAATCTTAGTCTGCTGCT | 865 | DB | 1808 | ATGTGTCGAGAGGAAAGACAGAGGAGGCTTGGCCCTAATATAAGAGTATAGTCA | 1867 |
| QY | 522 | TCATGCCATCGTGCCTGCTACATCAGTAGTGTGGTGGCTGGATCGTACGATACAGG | 581 | QY | 1542 | TTTTGGCGCTTCTAATGCTTGTGATGATTTCCGTGTCCACTGACGTGGGTGACGAGA | 1601 |
| DB | 866 | TCATTGCTATAGTGCAGAGTACATATCCAGATCAGTAGCTGATCATTTGACATGAAG | 925 | DB | 1868 | CTATGTTGATGCTAATGCTTGTGATGATGTTTGTGTACACTGTGCTGGTAACTAGCA | 1927 |
| QY | 582 | GCATTGCTATTCGCGCTGAGTTCACCTACTTCTGTGGGTGGCTCAGTGAAGCTG | 641 | QY | 1602 | ATGCTTACTCCAGTCCCTCCATGCTTCCACAACTGCAAGATGATGCCGCA | 1661 |
| DB | 926 | GCATTGCTATTCGCACTGCAATTTACATATATTTTGGGTAATACTGTAAATAACAG | 985 | DB | 1928 | ATGCTTACTCCAGCCCTAGTGTGTTATGGCTCATCAA---TCAGATGGAACAGAA | 1984 |
| QY | 642 | GATCCGTTCTGCTGGCGCAGCGCTTGTCTACTTCTACATGTTGTCGCGCTGGG | 701 | QY | 1662 | ACATTTTAGACGATTTTCAGAGAGGCTTACTTCTGGCTTTTCGAGAACACTGCCGATG | 1721 |
| DB | 986 | GCTGTGTTCTGCGCAATAGTTGTGTGCTGTATTTTATGTTTCCGCTGGG | 1045 | DB | 1985 | ATATCTGGATGATTTTCAGAGAGCATACTTCTGGCTAAGGAGAGATACAGATGAACATG | 2044 |
| QY | 702 | GTGGCTACGTTTATCATCAACTGATACCCCTGCACTCTTGTGTAATGCTCATATGG | 761 | QY | 1722 | CTCGGTTATGCTTGGTGGATTACCGATACCGATAGCGGGAATGGCAACAGAACGA | 1781 |
| DB | 1046 | GTGGTTATGTTTATCAATTAATCTTATTCACATACATGTTTGTACTGTTACTGATGC | 1105 | DB | 2045 | CCGCTGTATGCTCCTGGTGGGATTA---GGCTACCAAAATTCAGGAATGGCTTAATCGAACCA | 2104 |
| QY | 762 | GCAGTACTCGCGCGCTGCTGACAGCTACAGCAGCTTCTACATCTCGGAGCTGCTGT | 821 | QY | 1782 | CGCTAGTGGATAATAATACGTGGAACAATAGTCACATAGCGCTGGTGGCAAGGCAATGT | 1841 |
| DB | 1106 | AGAGATACAGCAAGAGTCTCATAGCTTACAGCACTTTTACATTTGGGTTTAATAC | 1165 | DB | 2105 | CTTTAGTAGACAAATACGTGGAAACAATAGCCACATAGCTTTGGTTGGAAGCAATGT | 2164 |
| QY | 822 | TCTCCATGCAAGATCCCTTCTGCTGGATTCCAAACGATACGCAACAGTGAACATGGCTG | 881 | QY | 1842 | CTTCAACCGGAGAGAGTCTCTACGAAATTAATGACATCTCTGACGTGGAGTACGTTTGG | 1901 |
| DB | 1166 | TATCAATGCAATTCCTTGTGGGATTCAGGCAATACGCAAGTGAACATATGGCAG | 1225 | DB | 2165 | CTTCAAAATGAACCGCTGCTTATGAATAATGAAGAGCTTGGATGTAGATTTATGTAATA | 2224 |
| QY | 882 | CGTGGAGGTTTGTGCTCTTATGGCGTGGCCACCTTGGCCATTTGAGTCCGCTGC | 941 | QY | 1902 | TGATCTTTTGGGTTGCTCGGCTATTCTGGCGATGATATCAACAAGTTCTCTGTGGATGG | 1961 |
| DB | 1226 | CTCAGGGGCTTTGGCCCTGCTGCAAGCTATGCCCTCTCTGATCTTACAGAGCAAC | 1285 | DB | 2225 | TAATATTTTGGGTTGTAATTTGGATATCTGCTGATGACATCAACAATACTCTCGATGG | 2284 |
| QY | 942 | TGTCGCAACAGATTCGGGAAGCTGTTCACTCGCGCGGATTTGCTGGGCGTTGGCG | 1001 | QY | 1962 | TCCGAATTTGTGAGGAGAGCATCCCAAGGACATTAAGGAACGATTTACTTTACCGACC | 2021 |
| DB | 1286 | TAAGTAGCAGGATTTCCAGACACTGTTTCTCGGTGATCGTTGGCAGCTGGAGCTG | 1345 | DB | 2285 | TTCGAATAGCAGAGGAGAGACATCCAAAGATATCAGGGAAAGTGACTACTTCCACC--AC | 2342 |
| QY | 1002 | TCCTTGTGGCGTGGTGTCTACCATGCTGGCGGTTTGGCGCGCTGGAGTGGAGCTG | 1061 | QY | 2022 | CGGCTGAATTCAGGTTAGATGCCAGGCTGCTCGGCGCTTCTCAACTGCTTATCTATCA | 2081 |
| DB | 1346 | TATTTCTCACTGTCACTATTTGACTTTATACAGGCTACATGTCTCCATGAGTGAAGA | 1405 | DB | 2343 | AAGGAGATTTCTGTTAGACAAAGCTGGATCTCCAACCTTGTCTCAATTGGCTCATGATA | 2402 |
| QY | 1062 | TCTACTCGCTGTGGGATCTGGCTTACGCCCAAGATCCACATTTCCCATTTGCAATCCGCTG | 1121 | QY | 2082 | AATTAAGCTACTCAGATTTCCGGGAATTTGAAGTTGGAATACAGAGTTCATCTGGATATG | 2141 |
| DB | 1406 | TTTACTCACTTTGGGACACAGGATATGGGAATTCAGAAATCCATATCCCATTCATCTGCTG | 1465 | DB | 2403 | AGATGTCAATTAACCGGTTTGGTGAATGCAGCTTGAATTTCCGAACACCACTGGGATTTG | 2462 |
| | | | | QY | 2142 | ATCGCACATGACCGCTCATTTGGGAATGAAGACTTTCGATCTGACCTCCTGGAGGAGG | 2201 |
| | | | | DB | 2463 | ACCGACACGTAATGCTGAAATTTGGGAATGAAGATTAATAATTCAGCATTTGGAAGAAG | 2522 |
| | | | | QY | 2202 | CCTACACACAGAACACTGGCTTGTTCGCTATATAGGGTGAAGAGCGCATGAGTTCA | 2261 |

Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
 Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216]

COMMENT
 Please visit our web site (http://genome.gsc.riken.jp/) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAAGATCCAGACTCTTTTITTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTTAATTAATTAATCCGCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
 Host: SOLR.

FEATURES
Location/Qualifiers
 1..2669
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:2610524N02"
 /db_xref="taxon:10090"
 /clone="2610524N02"
 /tissue="whole body"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days embryo"
 36..72507
 /note="hypothetical Oligosaccharyl transferase (OTase) STH3 subunit containing protein (InterPro|IPR003674, evidence: InterPro)
 putative"
 2654..2659
 /note="putative"
 2669
 /note="putative"

misc_feature
 polyA_signal
 2654..2659
 /note="putative"
 polyA_site
 2669
 /note="putative"

ORIGIN
 Query Match 39.1%; Score 944; DB 3; Length 2669;
 Best Local Similarity 64.9%; Pred. No. 2.1e-245;
 Matches 1498; Conservative 0; Mismatches 790; Indels 20; Gaps 6;
 QY 102 AGGTGGCTGGCTACAGCAGCTAATCACCTCGCCATCTGCTAATCGCTCGCTGGCCGG 161
 Db 217 AGCGGCGGGTGGCAGTGGTCTCTCTCTTACCATCTCTTCTGGCTGGCTGGCCGG 276
 QY 162 GATTTTCTCTCGCTCTTTCGGCGTCATCCGTTTCGAGTCGATATCCATGATTTGATPC 221
 Db 277 GCTTCAGCTCGCGCTCTTTCGCGGTCATCCGGTTCGAGAGCATCATCCAGAGTTCGACC 336
 QY 222 CGTGGTTCAACTACCGGGCCACCGCTACATCGTCGAGATCGTTGGTCAACTTCCTCA 281
 Db 337 CGTGGTTCACTATAGATCAACACATCATCTTGCATCTCATGATTCATGAGTTCTAA 396
 QY 282 ACTGTTTCAGGAGCGCGCATGGATTCGCTCGCAGGATTCGGCGGTACCGTCTATC 341
 Db 397 ATTGGTTTGATGAAGAGCATGGTACCACTGGGAAGAAATAGTGGTGGCCCGTTTACC 456
 QY 342 CGGCGCTGATGATTAGCTCCGCGCAATTCATTGCTCTGCTGCTGCTGCTGCTGCTGCT 497

457 CAGGGTTGATGATTAACAGCTGGCTTATTCATTGGAT-TTAAATACATTGAACATAACAG 515
 402 TCCATATTCGTGACATCTGGGTGTTCTGGCGCGGATCTTCAGTGGCCTGACCTCATCT 461
 516 TTCAATAAGAGATGTGTGTATTTCTGACCAACTTTTAGGGCTTACATCATAT 575
 462 CCACCTACCTGTGACCAAGAGCTGTGTGTCGCGGCGCGGCTTTCGCGCCAGCT 521
 576 CTAGCTTCTGCTAACTAGAGAACTGTGGAACAAGAGAGAGACTTCTAGCTGCTTGT 635
 522 TCATCGGCATCGCTGCTGCTGATCATCATGATAGTTCGTGCTGATCATGATACAGAG 581
 636 TCATGCTATCTGACAGAGTACATATCTCGTCACTGCTGCTGCTGCTGCTGCTGCT 695
 582 GCATGCTATTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
 696 GCATGCTATTCGCGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 755
 642 GATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701
 756 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815
 702 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
 816 GAGGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
 762 GCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 821
 876 AGAGTACAGCAAGAGAGTCTACATGATATAGCACTTCTACATGCTGCTGCTGCTGCT 935
 822 TCTCATGAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 881
 936 TATCATGAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 995
 882 CGCTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 940
 996 CTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1055
 941 CTGCTGGGCAACAGATTCGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000
 1056 TTGACAAAACAGGAGTTCAGAGCCCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1115
 1001 GCTCTTGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1060
 1116 GTGTTCTTAGTGTCTATCTGACATACAGAGTATATGACCATGAGTGGCAGG 1175
 1061 TTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1120
 1176 TTTTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
 1121 TCGAGATCAGCCGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1180
 1236 TCTGACATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1295
 1181 GCCTTCCAGTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1240
 1296 ACCTTCCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1355
 1241 GTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1300
 1356 GCTCTGATGGATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
 1301 CTCAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1360
 1416 CTGACCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1475
 1361 CTGCAAGAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1420
 1476 TTGGGGG-----GTGACATGAAGAGGAAACCCACCTGTGGAG-----CACAGCAGT 1523
 1421 GAAGCTGAGGATTCATATGAGAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1480

1524 GATGAGGATGACAAAGAAACCCAGGAAACTTGTATGACAGGACGTAAGTGAAGGAG 1583
 1481 CGTACTTACGATGATGCGCCAGCAGGATATCTGGCTCAGCTCAAACCTGAAGAGTATTGTT 1540
 1584 CATGTGACAGAGCAAGAAACCTGAGAGGGCTTGGGCCCAACATCAAAAGCAITGTG 1643
 1541 ATTTTGGCGCTTCTAATGCTGTGTATGATGTTGGTCTCACTGACAGTGGGTGACGAG 1600
 1644 ACCATGCTGATCTCATGCTCTGATGATGTTCGGGTCCACTGCTGCTGCTGCTGCTGCT 1703
 1601 AATGCTTACTCCAGTCCCTCCATGCTTGGCTTCCACACAGTCAAGATGATCCCGC 1660
 1704 AAGCTTACTCCAGTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1760
 1661 AATTTTATGAGATTTTACAGAGGCTTACTACTGCTTTCGAGAACACTGCGGATGAT 1720
 1761 AATATATTAGATGATTTTACAGAGGCTTACTTTTGGCTGAGACAAACACCGATGAAC 1820
 1721 GCTCGCTTATGCTTGTGGGATTTACGATACAGATAGCGGAATGGCAAAACAGAACG 1780
 1821 GCCCGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1880
 1781 ACCTTATGCTGATTAATATGCTGGAACAAATAGTACATAGCGCTGCTGCTGCTGCTGCT 1840
 1881 ACTCTGCTGATTAACACACCTGGAACCAACACGACCTGCTGCTGCTGCTGCTGCTGCT 1940
 1841 TCTTCAACCGAGGAGAGTCTTACGAAATATGACATCTTGTGACGTGGACTAGCTTTG 1900
 1941 TCTTCAATGAACCGCGCTTATTAATATCATGAGTCCCTTGTATGCTGATTTGCTG 2000
 1901 GTGATCTTTGGCGGTGTGATGCTGCTTATCTGCGATGATATCAACAAGTTCCTGCTG 1960
 2001 GTTATTTTTCGAGAGGATGATTTGCTTATTCGCGGACGATATCAACAAGTTCCTGCT 2060
 1961 GTCCGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTAC 2020
 2061 GTGAGATAGCTGAGGAGGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTAC 2120
 2021 CGCGGTGATTTGAGGTGATGCTGCGAGTGTCTGCGGCTGCTCAACTGCTTATGCTAC 2080
 2121 CAGGAGAGTTCGAGTACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2180
 2081 AATTAAGCTACTACAGATTTGGGGAATGAAATGAAATGAAATGAAATGAAATGAAAT 2140
 2181 AAAATGCTACTACAGATTTGGGGAATGAAATGAAATGAAATGAAATGAAATGAAAT 2240
 2141 GATCGCACACGTAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2200
 2241 GACCGAACACGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2300
 2201 GCTTACACACAGAAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2260
 2301 GCTTTTACATCAGACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2360
 2261 AATAGACATCATGAAACCAAGAGAGAGATTCCTCGACAACTTCATTTCCAGA 2320
 2361 GAGACACTAGGTCTCAAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2420
 2321 AAGAACTTAAGCGTGGCAGG---GGCTACATAGAAACCGACCGGTTGTTGTTAAGGGA 2377
 2421 AAGAGACTTAAAGGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2480
 2378 AAGAGAACCTTGAATTAACCCCAAAAA 2405
 2481 AAGAGAACCTTGAAGAGACTGTTTAAA 2508

RESULT 4
 CNS017ZX
 LOCUS

836 bp DNA linear GSS 26-JUL-1999
 Drosophila melanogaster genome survey sequence T7 end of BAC
 BACNS7M18 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

Margaret Fuller. Sized fractionated cDNAs were directly ligated into pOTB7. Plasmid cDNA library."

```

ORIGIN
Query Match 33.4%; Score 807.4; DB 2; Length 810;
Best Local Similarity 99.8%; Pred. No. 2.6e-208;
Matches 808; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 23 GCATTTTCAGATCGGTTAATATTTTCGAGTTACTGGTGAATTTGGACATGAATCGGAGC 82
DB 1 GCATTTTCAGATCGGTTAATATTTTCGAGTTACTGGTGAATTTGGACATGAATCGGAGC 60

QY 83 CCGAAGATCTGAACAGCAAGTGGCTGGCTACAGAGCCTAATACCTTCGCCATCTCG 142
DB 61 CCGAAGATCTGAACAGCAAGTGGCTGGCTACAGAGCCTAATACCTTCGCCATCTCG 120

QY 143 CTAATCGGCTGGCTGGCGGATTTTCTCTCGCTCTTTCGCGTCACTCGGTTTCGAGTCG 202
DB 121 CTAATCGGCTGGCTGGCGGATTTTCTCTCGCTCTTTCGCGTCACTCGGTTTCGAGTCG 180

QY 203 ATTATCCATGATTTGATTCGCTGTTCAACTACCGGCGCAACCGCTACATGTTGCAAT 262
DB 181 ATTATCCATGATTTGATTCGCTGTTCAACTACCGGCGCAACCGCTACATGTTGCAAT 240

QY 263 GGTGGTACAACTTCTCAACTGGTTCGAGGCGCGCATGTTATCGGTCGCGAGGATT 322
DB 241 GGTGGTACAACTTCTCAACTGGTTCGAGGCGCGCATGTTATCGGTCGCGAGGATT 300

QY 323 GTGGCGGTACGTTATTCGCGGCTGATGATTTACGTCGCGCGGAATTCATTGGCTGCTG 382
DB 301 GTGGCGGTACGTTATTCGCGGCTGATGATTTACGTCGCGCGGAATTCATTGGCTGCTG 360

QY 383 CACGTACTCAACATACCGGTCATATTCGTGACATCTGCTGTTCTCGCGCGCATCTTC 442
DB 361 CACGTACTCAACATACCGGTCATATTCGTGACATCTGCTGTTCTCGCGCGCATCTTC 420

QY 443 AGTGCGCTGACCTGATCTCACTACTGCTGACAGGAGCTGTGGTCGCGGCGGC 502
DB 421 AGTGCGCTGACCTGATCTCACTACTGCTGACAGGAGCTGTGGTCGCGGCGGC 480

QY 503 GCGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCCTGGCTACATCACTAGTGTGCTGGCT 562
DB 481 GCGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCCTGGCTACATCACTAGTGTGCTGGCT 540

QY 563 GATCGTACGATACAGGAGGATTTGCCATATTTGCGCTGCGCTGACCTACTTCTGCTGG 622
DB 541 GATCGTACGATACAGGAGGATTTGCCATATTTGCGCTGCGCTGACCTACTTCTGCTGG 600

QY 623 GTGCGCTCAGTGAAGACTGGATCGGTGTTCTGTGTCGCGCGCAGCGCTTGTCTACTTTC 682
DB 601 GTGCGCTCAGTGAAGACTGGATCGGTGTTCTGTGTCGCGCGCAGCGCTTGTCTACTTTC 660

QY 683 TACATGGTTCGCGCTGGGTGGCTAGTGTTCATCATCACTGATACCCCTGACGCTC 742
DB 661 TACATGGTTCGCGCTGGGTGGCTAGTGTTCATCATCACTGATACCCCTGACGCTC 720

QY 743 TTCGTACTGCTCAATATGGGAGGACTCTCGCGGCTGCTGTGACCACTACAGCACTTC 802
DB 721 TTCGTACTGCTCAATATGGGAGGACTCTCGCGGCTGCTGTGACCACTACAGCACTTC 780

QY 803 TACATCTCGGAGCTGCTGTTCTCCATGCAG 832
DB 781 TACATCTCGGAGCTGCTGTTCTCCATGCAG 810

```

RESULT 6

CK657451
 LOCUS LP23759 Sprime LP Drosophila melanogaster larval-early pupal pOT2
 DEFINITION Drosophila melanogaster cDNA clone LP23759 5, mRNA sequence.
 ACCESSION CK657451
 VERSION CK657451.1 GI:41400976
 KEYWORDS EST.

SOURCE ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G.M.
 BDGP/HMM Drosophila EST Project
 Unpublished (2001)

TITLE JOURNAL COMMENT

Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 Plate: LP.237 row: E column: 11
 High quality sequence stop: 644.

FEATURES source

1..715
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="LP23759"
 /sex="male and female"
 /dev_stage="larvae-pupae"
 /lab_host="DHS-alpha"
 /clone_lib="LP Drosophila melanogaster larval-early pupal
 pOT2"
 /note="Organ: whole body; Vector: pOT2; Site 1: EcoRI;
 Site 2: XhoI; Sized fractionated cDNAs were directly
 ligated into pOT2. Plasmid cDNA library."

ORIGIN

```

Query Match 29.4%; Score 711.8; DB 7; Length 715;
Best Local Similarity 99.7%; Pred. No. 2.8e-182;
Matches 713; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 AATGTCTCGTTCGATTCAGATCGGTTAATATTTTCGAGTTACTGGTGGAAATGGGACA 71
DB 1 AATGTCTCGTTCGATTCAGATCGGTTAATATTTTCGAGTTACTGGTGGAAATGGGACA 60

QY 72 TGAATCGGACGCGGAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCCTAATCACT 131
DB 61 TGAATCGGACGCGGAGATGCTGAACAGCAAGTGGCTGGCTACAGAGCCTAATCACT 120

QY 132 TCGCCATCTGCTAATTCGCTGGCTGGCGGATTTTCTCTCGGCTCTTTCGCGCTATCC 191
DB 121 TCGCCATCTGCTAATTCGCTGGCTGGCGGATTTTCTCTCGGCTCTTTCGCGCTATCC 180

QY 192 GTTTCAGTCGATTATCCATGATTTGATTCGTTGTTCAACTACCGGCGCACCCCTACA 251
DB 181 GTTTCAGTCGATTATCCATGATTTGATTCGTTGTTCAACTACCGGCGCACCCCTACA 240

QY 252 TGGTCGAGAAATGGTGGTAACTTCTCACTGGTTTCGAGAGCGCGCATGGTATCCGC 311
DB 241 TGGTCGAGAAATGGTGGTAACTTCTCACTGGTTTCGAGAGCGCGCATGGTATCCGC 300

QY 312 TCGGCAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGGAATCC 371
DB 301 TCGGCAGGATTTGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGGAATCC 360

QY 372 ATTGGCTGTGACGTTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTCTGG 431
DB 361 ATTGGCTGTGACGTTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTCTGG 420

QY 432 CGCGGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACTCTGTGACCAAGAGCTGTGG 491
DB 421 CGCGGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACTCTGTGACCAAGAGCTGTGG 480

QY 492 CGCGGCGCGCGGCTCTTTCGCGCGGAGTTTCAACGCGGCTGTGCTGGCTACATCAGTA 551
DB 481 CGCGGCGCGCGGCTCTTTCGCGCGGAGTTTCAACGCGGCTGTGCTGGCTACATCAGTA 551

```


estimated-cyto:96B10-96C1: 04/09/2001
 Plate: At.194 row: A column: 6
 High quality sequence stop: 602.
 Location/Qualifiers
 1. 701
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="At19406"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_hosts="Plates At.10-At.120: DHS-alpha. Plates
 AT.121-AT.319: DHS-alpha Tona"
 /clone_lib="AT Drosophila melanogaster adult testes POTB7"
 /note="Organ: ADULT testes; Vector: pOPB7; Site: 1: EcoRI;
 Site_2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into POTB7. Plasmid cDNA library."

FEATURES
 source

ORIGIN
 Query Match 28.9%; Score 699.4; DB 2; Length 701;
 Best Local Similarity 99.9%; Pred. No. 6.7e-179; Indels 0; Gaps 0;
 Matches 700; Conservative 0; Mismatches 1;
 QY 739 CGTCTGCTACTGCTATTATGGCAGGTACTCGCGCGTCTGCTGACGAGTACAGCAC 798
 Db 1 CGTCTGCTACTGCTATTATGGCAGGTACTCGCGCGTCTGCTGACGAGTACAGCAC 60
 QY 799 CTTCTACATCTGGAGTCTGTTCTCCATGACAGATCCCTTCGCGGATTCACACCGAT 858
 Db 61 CTTCTACATCTGGAGTCTGTTCTCCATGACAGATCCCTTCGCGGATTCACACCGAT 120
 QY 859 ACGCACCAGTGAACACATGCTGCGCTGGGAGTGTGTTGCTCTTATGCGCGTGGCCAC 918
 Db 121 ACGCACCAGTGAACACATGCTGCGCTGGGAGTGTGTTGCTCTTATGCGCGTGGCCAC 180
 QY 919 CTTGCGCATTTGAGTCCGCTGCTGCGCAGACAGTTCGGAAGCTTTATGCTGCGG 978
 Db 181 CTTGCGCATTTGAGTCCGCTGCTGCGCAGACAGTTCGGAAGCTTTATGCTGCGG 240
 QY 979 CGGATTGCTGCGCGTTCGCTTTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1038
 Db 241 CGGATTGCTGCGCGTTCGCTTTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 QY 1039 TGTGCCCCGTTGAGTGGAGTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1098
 Db 301 TGTGCCCCGTTGAGTGGAGTGGAGTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 QY 1099 CATTCCCATATTCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1158
 Db 361 CATTCCCATATTCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 1159 TGATCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1218
 Db 421 TGATCTGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 QY 1219 CAACGACGAGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278
 Db 481 CAACGACGAGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 1279 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
 Db 541 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 QY 1339 TTCGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1398
 Db 601 TTCGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 QY 1399 AAGCGCAGCCAGGAGTGAAGCTGAGGATTCATTG 1439
 Db 661 AAGCGCAGCCAGGAGTGAAGCTGAGGATTCATTG 701

RESULT 9
 BG640860
 LOCUS
 DEFINITION

690 bp mRNA linear EST 23-APR-2001
 SD12024.5prime SD Drosophila melanogaster Schneider L2 cell culture
 POT2 Drosophila melanogaster cDNA clone SD12024.5 similar to
 CatStt3: Fban0007748 'enzyme' located on: 3R 96B16-96B17;
 04/13/2001, mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BG640860.1 GI:13772786
 EST.
 Drosophila melanogaster (fruit fly)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 AUTHORS

1 (bases 1 to 690)
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S. and Rubin, G.M.
 BDGP/HMI Drosophila EST Project
 Unpublished (2001)

TITLE
 JOURNAL
 COMMENT

Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AB003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 04/13/2001 hit P element 1(3)j2D9:
 1(3)j2D9 A0026308 inserted at base 292 5' end of P element Inverse
 PCR: 03/20/2001 row: B column: 12
 Plate: SD:120 row: B column: 12
 High quality sequence stop: 535.

FEATURES
 source

1. 690
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="SD12024"
 /lab_host="DHS-alpha"
 /clone_lib="SD Drosophila melanogaster Schneider L2 cell
 culture POT2"
 /note="Vector: POT2; Site: 1: EcoRI; Site 2: XhoI; Sized
 fractionated cDNAs were directly ligated into POT2.
 Plasmid cDNA library."

ORIGIN

Query Match 28.5%; Score 688.4; DB 4; Length 690;
 Best Local Similarity 99.9%; Pred. No. 6.6e-176; Indels 0; Gaps 0;
 Matches 689; Conservative 0; Mismatches 1;
 QY 21 TTGCATTTCAGATCGGTATATATTTTCGAGTTACTGGCTGGAATTGGGACATGAATCGGA 80
 Db 1 TTGCATTTCAGATCGGTATATATTTTCGAGTTACTGGCTGGAATTGGGACATGAATCGGA 60
 QY 81 CGCCGAGATGCTGAACAGCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 140
 Db 61 CGCCGAGATGCTGAACAGCAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 QY 141 TGCTAATCCCTGGGTGGCGGATTTTCTCTCGCTCTTCGCGCTCATCGCTTTCGAGT 200
 Db 121 TGCTAATCCCTGGGTGGCGGATTTTCTCTCGCTCTTCGCGCTCATCGCTTTCGAGT 180
 QY 201 CGATTATCATGAGTTGATTCGTTCACTACCGGCGCCACCGCTTACATGCTGTCGAGA 260
 Db 181 CGATTATCATGAGTTGATTCGTTCACTACCGGCGCCACCGCTTACATGCTGTCGAGA 240
 QY 261 ATGGTTGGTACAACTTCCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 320
 Db 241 ATGGTTGGTACAACTTCCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 QY 321 TTGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGGGAATCCATTGGCTGC 380

BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic AE003750: arm:3R [20671983,20899380]
estimated-cyto:96B10-96C1: 05/18/2001
Plate: GM.263 row: D column: 2
High quality sequence stop: 585.
Location/Qualifiers

FEATURES

1..694
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GM26338"
/sex="female"
/dev_stage="newly eclosed females: germarium-stage 6"
/lab_host="XUI Blue"
/note="Organ: ovary; Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."
Location/Qualifiers

ORIGIN

Query Match 28.2%; Score 690.4; DB 4; Length 694;
Best Local Similarity 98.6%; Pred. No. 1e-173;
Matches 684; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1497 CCCAGCAGGATACCTGGCGTCCAGCTCCAACTGAGAGTATTGTTATTTGGCGCTTAA 1556
DB 1 CCCAGCAGGATACCTGGCGTCCAGCTCCAACTGAGAGTATTGTTATTTGGCGCTTAA 60

QY 1557 TCGTGTGATGATGTTGCTGTCCATGCACTGACGTGGGTGACAGCAATGCTACTCAGTC 1616
DB 61 TCGTGTGATGATGTTGCTGTCCATGCACTGACGTGGGTGACAGCAATGCTACTCAGTC 120

QY 1617 CCTCATTGCTTGGCTTCCAACTCAAGTCAAGTGGATCCCGCAACATTTAGACGATT 1676
DB 121 CCTCATTGCTTGGCTTCCAACTCAAGTCAAGTGGATCCCGCAACATTTAGACGATT 180

QY 1677 TCAGAGAGCTTACTACTGGCTTTCGAGAACACTGCCGATGATGCTCGGTATGCTT 1736
DB 181 TCAGAGAGCTTACTACTGGCTTTCGAGAACACTGCCGATGATGCTCGGTATGCTT 240

QY 1737 GGTGGGATTACGGATACAGATACGGGAATGCAACAGAGAGCTAGTGGATAATA 1796
DB 241 GGTGGGATTACGGATACAGATACGGGAATGCAACAGAGAGCTAGTGGATAATA 300

QY 1797 ATAGTGGAAATAGTACATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGA 1856
DB 301 ATAGTGGAAATAGTACATAGCGCTGTTGGCAAGCAATGCTTCAACCGAGGAGA 360

QY 1857 AGTCCTACGAAATATGACATCTCTTGACGTGACATACGTTTGGTGATCTTTGGCGGTG 1916
DB 361 AGTCCTACGAAATATGACATCTCTTGACGTGACATACGTTTGGTGATCTTTGGCGGTG 420

QY 1917 TGATCGGTATTCTGGCGATGATATCAAGTTCCTGTGGATGTCGGAATGCTGAGG 1976
DB 421 TGATCGGTATTCTGGCGATGATATCAAGTTCCTGTGGATGTCGGAATGCTGAGG 480

QY 1977 GAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCGACCGCGTGAATTCAGGG 2036
DB 481 GAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTTTACCGACCGCGTGAATTCAGGG 540

QY 2037 TAGATGCGGAAGTGTCTCCGCCCTGCTCAACTGCTTATGTATACAAATTAAGCTACTACA 2096
DB 541 TAGATGCGGAAGTGTCTCCGCCCTGCTCAACTGCTTATGTATACAAATTAAGCTACTACA 600

QY 2097 GATTTCGGGAATGAAGTTGGAATACAGAGGTCCTATCTGGATATGATCCACACCTAACG 2156
DB 601 GATTTCGGGAATGAAGTTGGAATACAGAGGTCCTATCTGGATATGATCCACACCTAACG 660

QY 2157 CCGTCATGGGAATTAAGGACTTCGATCTGACCTA 2190

DB

661 CCGTCATGGGAATTAAGGACTTCGATCTGACCTA 694

RESULT 12

CK657649

LOCUS

DEFINITION

CK657649

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 679)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S. and Rubin, G.M.

BDGP/HMM Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

Plate: LP.241 row: B column: 7

High quality sequence stop: 670.

FEATURES

Location/Qualifiers

1..679

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP24119"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DH5-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal

pOT2"

/note="Organ: whole body; Vector: pOT2; Site 1: EcoRI;

Site 2: XhoI; Sized fractionated cDNAs were directly

ligated into pOT2. Plasmid cDNA library.

ORIGIN

Query Match 28.0%; Score 677.4; DB 7; Length 679;
Best Local Similarity 99.9%; Pred. No. 6.6e-173;
Matches 678; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 18 TCCTTCATTCAGATCGGTATATATTTTCGAGTTACTGCTGGAATTGGACATGAATC 77
DB 1 TCCTTCATTCAGATCGGTATATATTTTCGAGTTACTGCTGGAATTGGACATGAATC 60

QY 78 GAGCGCGAAGATGCTGACAGCAGGTGGTGGCTACAGCAGCCTAATCACCTTCGCA 137
DB 61 GAGCGCGAAGATGCTGACAGCAGGTGGTGGCTACAGCAGCCTAATCACCTTCGCA 120

QY 138 TCCTGCTAATCCCTGGGTGCGGATTTCTCTTCGCTCTTCGCGTCAATCCGTTTCG 197
DB 121 TCCTGCTAATCCCTGGGTGCGGATTTCTCTTCGCTCTTCGCGTCAATCCGTTTCG 180

QY 198 AGTCGATTCATTCAGTTGATCCGTGTTCACTACCGGCGCACCGCTACATGTTGC 257
DB 181 AGTCGATTCATTCAGTTGATCCGTGTTCACTACCGGCGCACCGCTACATGTTGC 240

QY 258 AGAATGGTTGGTACAACTTCCTCAACTGTTTCGACGAGCGCGCATGGTATCCGCTCGCA 317
DB 241 AGAATGGTTGGTACAACTTCCTCAACTGTTTCGACGAGCGCGCATGGTATCCGCTCGCA 300

QY 318 GGATTCGGCGGTACCTCTATCCGCTGATGATACGTCGGCGGATCCATTCGC 377

Db 301 GGATTGTGGCGGTACCGTCTATCCGCGCTGATGATTACGTCGGCGGAATCCATTGGC 360
 QY 378 TGTCTCAGTACTCAACATACCGTCCATATTCGTGACATCTCGTCTTCTCGCGCGA 437
 Db 361 TGTCTCAGTACTCAACATACCGTCCATATTCGTGACATCTCGTCTTCTCGCGCGA 420
 QY 438 TCTTCACTGCTGCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCT 497
 Db 421 TCTTCACTGCTGCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCT 480
 QY 498 GCGCGGCTCTTCCGCGCGCTCACTGCGCATCTGCTGCGCATCTCACTGCTGCTGCT 557
 Db 481 GCGCGGCTCTTCCGCGCGCTCACTGCGCATCTGCTGCGCATCTCACTGCTGCTGCT 540
 QY 558 TGGCTGATCGTACGATACGAGGCAATTCGATATTCGCGCTGCGCTTCACTTCTTC 617
 Db 541 TGGCTGATCGTACGATACGAGGCAATTCGATATTCGCGCTTCACTTCTTC 600
 QY 618 TGTGGTGGCTCAGTGAGACTGGATCCGTTCTGTCGGCGCGAGCGCTTGTCTCT 677
 Db 601 TGTGGTGGCTCAGTGAGACTGGATCCGTTCTGTCGGCGCGAGCGCTTGTCTCT 660
 QY 678 ACTTCTACATGCTGCTCCG 696
 Db 661 ACTTCTACATGCTGCTCCG 679

RESULT 13

CK659064
 LOCUS
 DEFINITION LP16523.5prime LP Drosophila melanogaster larval-early pupal pot2
 Drosophila melanogaster cDNA clone LP16523 5, mRNA sequence.
 CK659064
 CK659064.1 GI:41402589
 EST.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 682)
 Harvey D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
 Lewis, S., and Rubin, G.M.
 BDGP/HMI Drosophila EST Project
 Unpublished (2001)
 Contact: Stapleton, M.
 BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
 Plate: LP165 row: B column: 11
 High quality sequence stop: 470.
 Location/Qualifiers

FEATURES
source

1..682
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="LP16523"
 /sex="male and female"
 /dev_stage="larvae-pupae"
 /lab_host="DHS-alpha"
 /clone_lib="LP Drosophila melanogaster larval-early pupal
 pot2"
 /note="Organ: whole body; Vector: pot2; Site 1: EcoRI;
 Site 2: XhoI; Sized fractionated cDNAs were directly
 ligated into pot2. Plasmid cDNA library."

ORIGIN

Query Match 28.0%; Score 676.2; DB 7; Length 682;
 Best Local Similarity 99.6%; Pred. No. 1.4e-172;
 Matches 678; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 15 GTGCGTTGCATTTTCAGATCGTTTATATTTTCGAGTTACTGTGCTGGAATTCGACATGA 74
 Db 2 GTGCGTTGCATTTTCAGATCGTTTATATTTTCGAGTTACTGTGCTGGAATTCGACATGA 61
 QY 75 ATCGGACCCGCAAGATGCTGAACAGAGGTGGCTGGCTACAGAGCCTAATCACTTCG 134
 Db 62 ATCGGACCCGCAAGATGCTGAACAGAGGTGGCTGGCTACAGAGCCTAATCACTTCG 121
 QY 135 CCATCTCTCTAATCGCTGGCTGGCGGATTTTCTCTCGCCCTCTTCGCGCTCATCGTT 194
 Db 122 CCATCTCTCTAATCGCTGGCTGGCGGATTTTCTCTCGCCCTCTTCGCGCTCATCGTT 181
 QY 195 TCGAGTGCATTTTCAATGAGTTTGAATCGTGGTTTCAATACCGGCGCAACGCTTACATGG 254
 Db 182 TCGAGTGCATTTTCAATGAGTTTGAATCGTGGTTTCAATACCGGCGCAACGCTTACATGG 241
 QY 255 TCGAGATGGTTGTGTAACAATCTTCAACTGGTTTCGACGAGCGCATGTATCCGCTCG 314
 Db 242 TCGAGATGGTTGTGTAACAATCTTCAACTGGTTTCGACGAGCGCATGTATCCGCTCG 301
 QY 315 GCAGATGTGGCGGTACCGTCTATCCGCGCTGATGATTACGTCGCGGGAATCCATT 374
 Db 302 GCAGATGTGGCGGTACCGTCTATCCGCGCTGATGATTACGTCGCGGGAATCCATT 361
 QY 375 GCGTCTGACGATCTCAACATACCGGTCCATATTCGTGACATCTGCGTCTTCTCGCGC 434
 Db 362 GCGTCTGACGATCTCAACATACCGGCGCATATTCGTGACATCTGCGTCTTCTCGCGC 421
 QY 435 GCATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGTGTCG 494
 Db 422 GCATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGTGTCG 481
 QY 495 CGCGCGCGCGCTCTTTCGCGCGCGAGTTCATCGCATCTGCTGCTGCTGCTGCTGCT 554
 Db 482 CGCGCGCGCGCTCTTTCGCGCGCGAGTTCATCGCATCTGCTGCTGCTGCTGCTGCT 541
 QY 555 CGGTGGCTGATCGTACGATACGAGGCGATTCGATATTCGCGCTGCTGCTGCTGCTGCT 614
 Db 542 CGGTGGCTGATCGTACGATACGAGGCGATTCGATATTCGCGCTGCTGCTGCTGCTGCT 601
 QY 615 TCCTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674
 Db 602 GCCTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661
 QY 675 CCTACTTCTACATGCTGTCG 695
 Db 662 CCTACTTCTACATGCTGTCG 682

RESULT 14

CK659064

LOCUS

DEFINITION

CK659064

CK659064.1

GI:41402589

EST.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 682)

Harvey D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

Lewis, S., and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

Plate: LP165 row: B column: 11

High quality sequence stop: 470.

Location/Qualifiers

1..682

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LP16523"

/sex="male and female"

/dev_stage="larvae-pupae"

/lab_host="DHS-alpha"

/clone_lib="LP Drosophila melanogaster larval-early pupal
 pot2"

/note="Organ: whole body; Vector: pot2; Site 1: EcoRI;
 Site 2: XhoI; Sized fractionated cDNAs were directly
 ligated into pot2. Plasmid cDNA library."

One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 04/13/2001 hit P element 1(3)J2D9:
 1(3)J2D9 A0026308 inserted at base 292 5' end of P element Inverse
 PCR: 03/20/2001
 Plate: SD.124 row: D column: 12
 High quality sequence stop: 604.
 Location/Qualifiers
 1. 663
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="SD12448"
 /lab_hosts="DHS-alpha"
 /clone_lib="SD Drosophila melanogaster Schneider L2 cell
 culture pOT2"
 /note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized
 fractionated cDNAs were directly ligated into pOT2.
 Plasmid cDNA library."

FEATURES
 source
 Query Match 27.3%; Score 659.8; DB 4; Length 663;
 Best Local Similarity 99.7%; Pred. No. 4.1e-168;
 Matches 661; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORIGIN
 1 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 60
 1 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 60
 61 GAATTGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
 61 GAATTGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
 121 CCTAATCACTTCGCCATCTCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 180
 121 CCTAATCACTTCGCCATCTCTGCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 180
 181 CGCGTCTACCTTCGATTCGATTCATTCATGATTTGATTCGCTGTTCACTACCGGC 240
 181 CGCGTCTACCTTCGATTCGATTCATTCATGATTTGATTCGCTGTTCACTACCGGC 240
 241 CACCGCTACATCGCTGCAGATGTTGGTACAACTTCCTCAACTGTTTCGAGCGGC 300
 241 CACCGCTACATCGCTGCAGATGTTGGTACAACTTCCTCAACTGTTTCGAGCGGC 300
 301 ATGGTATCCGCTCGCGAGATGTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 360
 301 ATGGTATCCGCTCGCGAGATGTTGGCGGTACCGTCTATCCGCGCTGATGATACGTC 360
 361 CGCGGATTCATTCGCTGCTGACGCTACCACTACCGCTCCATTCGTCGATC 420
 361 CGCGGATTCATTCGCTGCTGACGCTACCACTACCGCTCCATTCGTCGATC 420
 421 CGTGTTCCTGCGCGGATTCCTTCACTGAGTTCGCTGACCTCCATTCACCTACGCA 480
 481 GGAGCTGTGTCGCGCGGCGCGCTTCCTTCGCGCGGCTTCATCGCCATTCGCTGG 540
 481 GGAGCTGTGTCGCGCGGCGCGCTTCCTTCGCGCGGCTTCATCGCCATTCGCTGG 540
 541 CTACATCAGTAGTTCGCTGGCTGATGATGATGATGATGATGATGATGATGATGAT 600
 541 CTACATCAGTAGTTCGCTGGCTGATGATGATGATGATGATGATGATGATGATGAT 600
 601 GCAGTTCACCTACCTTCCTTCGCTGGCTGCTCAGTGAAGACTGGATCGGTTCTGG 660
 601 GCAGTTCACCTACCTTCCTTCGCTGGCTGCTCAGTGAAGACTGGATCGGTTCTGG 660
 661 CGC 663

Db 661 CGC 663

RESULT 15
 BI357074
 LOCUS
 DEFINITION
 Drosophila melanogaster cDNA clone R43425 5 similar to Osetc3:
 FBan0007748 'enzyme' located on: 3R 96B16-96B17; 05/13/2001, mRNA
 sequence.
 BI357074
 BI357074.1 GI:15051528
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 666)
 AUTHORS
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Mista, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S.,
 Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
 Rubin, G.M.
 BDGP/HMI RE Drosophila EST Project
 Unpublished (2001)
 Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AE003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 05/13/2001 hit P element 1(3)J2D9:
 1(3)J2D9 A0026308 inserted at base 292 5' end of P element Inverse
 PCR: 05/13/2001
 Plate: RE.434 row: C column: 1
 High quality sequence stop: 563.
 Location/Qualifiers
 1. 666
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE43425"
 /sex="male and female"
 /dev_stages="0-24 hours mixed stage embryonic"
 /lab_hosts="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pflc-1"
 /note="Organ: embryo; Vector: pflc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

FEATURES
 source
 Query Match 27.2%; Score 658; DB 4; Length 666;
 Best Local Similarity 100.0%; Pred. No. 1.3e-167;
 Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
 1 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 60
 9 TCTAAGCGAAGATGTCGTTCGATTCAGATTCGGTTATTAATTTTCGAGTTACTGGCTG 68
 61 GAATTGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 120
 69 GAATTGGACATGAATCGAGCGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGCAG 128
 121 CCTAATCACTTCGCTGCTCTCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 180
 129 CCTAATCACTTCGCTGCTCTCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTT 188
 181 CGCGGTCATCCGTTTCGAGTTCGATTCATTCATGATTTGATTCGTTCACTACCGGC 240

Db |||||CCCCGTCATCCCGTTCGAGTCGATTCATGAGTTGATCCGTTGGTTCACTACCGGGC 248
Qy |||||CACCGCCTACATGGTGCAGAAATGGTTGGTACAACTTCTCAACTGGTTCCAGAGCGCGC 300
Db |||||CACCGCCTACATGGTGCAGAAATGGTTGGTACAACTTCTCAACTGGTTCCAGAGCGCGC 308
Qy |||||ATGGTATCCGCTCCGCGAGGATTCGGCGGTACCGTCTATCCCGGCTGATGATTAGTC 360
Db |||||ATGGTATCCGCTCCGCGAGGATTCGGCGGTACCGTCTATCCCGGCTGATGATTAGTC 368
Qy |||||CGGCGGAATCCATTCGCTGCTGCAGTACTCAACATACCGTCCATATTCGTGACATCTG 420
Db |||||CGGCGGAATCCATTCGCTGCTGCAGTACTCAACATACCGTCCATATTCGTGACATCTG 428
Qy |||||CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGACCAA 480
Db |||||CGTGTTCCTGGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGACCAA 488
Qy |||||GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGGCCAGCTTCATCGCCATCGTGCCTGG 540
Db |||||GGAGCTGTGGTCCGCGGGCGCGGCTCTTCGCGGCCAGCTTCATCGCCATCGTGCCTGG 548
Qy |||||CTACATCAGTACGTCGCTGGTGGATCGTACGATAACGAGGGCATTCGCATATTCGCCCT 600
Db |||||CTACATCAGTACGTCGCTGGTGGATCGTACGATAACGAGGGCATTCGCATATTCGCCCT 608
Qy |||||GCAGTTCACCTACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTTCTGGTGG 658
Db |||||GCAGTTCACCTACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTTCTGGTGG 666

Search completed: December 15, 2004, 11:09:49
Job time : 7356.2 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 8574.94 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRTPMLNSKVAGYSSLIT.....RRKGYYIRNPVVVKGKRLK 774

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cygn2_1/USPRO_pool/US10028384/runat_14122004_131645_6800/app_query.fasta.1.3740
-DB=GenEmbl -QMT=fastap -SUFFIX=Oligo.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384.scgn.1.20262 @runat_14122004_131645_6800 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pt.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 774 | 100.0 | 2417 | 3 | AF132552 |
| 2 | 774 | 100.0 | 2417 | 6 | AX799088 |
| 3 | 774 | 100.0 | 2699 | 6 | CQ589353 |
| 4 | 464 | 59.9 | 4922 | 6 | CQ589352 |

| | | | | | |
|-----|------|--------|----|----------|--------------------|
| 464 | 59.9 | 162921 | 3 | AC007853 | Drosophil |
| 464 | 59.9 | 181132 | 3 | AC008206 | AC008206 Drosophil |
| 464 | 59.9 | 227219 | 3 | AE003750 | AE003750 Drosophil |
| 390 | 50.4 | 75650 | 3 | AC018145 | AC018145 Drosophil |
| 208 | 26.9 | 2953 | 6 | CQ596819 | CQ596819 Sequence |
| 83 | 10.7 | 2785 | 6 | CQ596795 | CQ596795 Sequence |
| 47 | 6.1 | 54118 | 3 | U13019 | U13019 Caenorhabdi |
| 42 | 5.4 | 433 | 6 | AX340470 | AX340470 Sequence |
| 42 | 5.4 | 507 | 6 | AX886102 | AX886102 Sequence |
| 42 | 5.4 | 507 | 6 | BD025712 | BD025712 Sequence |
| 42 | 5.4 | 764 | 6 | AX136480 | AX136480 Sequence |
| 42 | 5.4 | 764 | 6 | BD123720 | BD123720 Sequence |
| 42 | 5.4 | 787 | 6 | AX869456 | AX869456 Sequence |
| 42 | 5.4 | 787 | 6 | BD149518 | BD149518 Primer fo |
| 42 | 5.4 | 957 | 9 | BC015880 | BC015880 Homo sapi |
| 42 | 5.4 | 1097 | 5 | AY437849 | AY437849 Ctenophar |
| 42 | 5.4 | 1734 | 10 | BC003206 | BC003206 Mus muscu |
| 42 | 5.4 | 2284 | 6 | AX882932 | AX882932 Sequence |
| 42 | 5.4 | 2284 | 6 | BD160013 | BD160013 Primer fo |
| 42 | 5.4 | 2284 | 9 | AK027789 | AK027789 Homo sapi |
| 42 | 5.4 | 2481 | 6 | AY799082 | AY799082 Sequence |
| 42 | 5.4 | 2481 | 9 | AY074880 | AY074880 Homo sapi |
| 42 | 5.4 | 2503 | 6 | AX017997 | AX017997 Sequence |
| 42 | 5.4 | 2503 | 6 | BD137303 | BD137303 Human nuc |
| 42 | 5.4 | 2508 | 10 | BC013054 | BC013054 Mus muscu |
| 42 | 5.4 | 2510 | 6 | AX136135 | AX136135 Sequence |
| 42 | 5.4 | 2510 | 6 | BD123520 | BD123520 Secretary |
| 42 | 5.4 | 2510 | 9 | AK075380 | AK075380 Homo sapi |
| 42 | 5.4 | 2546 | 6 | AX099510 | AX099510 Sequence |
| 42 | 5.4 | 2546 | 6 | BD363986 | BD363986 Secreted |
| 42 | 5.4 | 2620 | 5 | CR386955 | CR386955 Gallus ga |
| 42 | 5.4 | 2710 | 6 | AX799084 | AX799084 Sequence |
| 42 | 5.4 | 4236 | 10 | BC052433 | BC052433 Mus muscu |
| 42 | 5.4 | 349 | 6 | BD071886 | BD071886 Secreted |
| 41 | 5.3 | 349 | 6 | BD071886 | BD071886 Secreted |
| 39 | 5.2 | 518 | 11 | BY001913 | BY001913 S210P6075 |
| 40 | 5.2 | 584 | 11 | G89286 | G89286 S208P6573FC |
| 40 | 5.2 | 203492 | 2 | AC107826 | AC107826 Mus muscu |
| 40 | 5.2 | 227953 | 2 | AC134292 | AC134292 Rattus no |
| 40 | 5.2 | 285379 | 2 | AC113070 | AC113070 Mus muscu |
| 39 | 5.0 | 1040 | 6 | CQ723424 | CQ723424 Sequence |
| 39 | 5.0 | 170691 | 9 | AC104643 | AC104643 Homo sapi |

ALIGNMENTS

| | | | | | | |
|------------|------------|-------------------------|---------|-------------|--------|-----------------|
| RESULT 1 | AF132552 | Drosophila melanogaster | 2417 bp | mRNA | linear | INV 12-FEB-2003 |
| LOCUS | AF132552 | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| DEFINITION | AF132552 | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| ACCESSION | AF132552 | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| VERSION | AF132552.1 | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| KEYWORDS | FLI_CDNA. | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| SOURCE | FLI_CDNA. | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| ORGANISM | FLI_CDNA. | Drosophila melanogaster | GM01838 | full insert | CDNA. | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | | | |
| AUTHORS | 1 | (bases 1 to 2417) | | | | |
| TITLE | 1 | (bases 1 to 2417) | | | | |
| JOURNAL | 1 | (bases 1 to 2417) | | | | |
| REFERENCE | 1 | (bases 1 to 2417) | | </ | | |

Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(<http://fruitfly.berkeley.edu>) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES

Source

1. 2417
/organism="Drosophila' melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
1. 2417

| | | | |
|------------------------|---------|---------------|------|
| Alignment Scores: | 0 | Length: | 2417 |
| Pred. No.: | 774.00 | Matches: | 774 |
| Score: | 100.00% | Conservative: | 0 |
| Percent Similarity: | 100.00% | Mismatches: | 0 |
| Best Local Similarity: | 100.00% | Indels: | 0 |
| Query Match: | 100.00% | Gaps: | 0 |
| DB: | 3 | | |

US-10-028-384-8 (1-774) x AF132552 (1-2417)

| | | | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| 1 | Met | Asn | Arg | Thr | Pro | Leu | Met | Leu | Asn | Ser | Leu | Val | Ala | Gly | Tyr | Ser | Ser | Leu | Leu | Thr | 20 |
| 71 | ATG | AAT | CGG | CGC | GCG | AGA | TCT | GAA | CAG | CAA | GCT | GCT | GCT | GCT | GCT | GCT | GCT | GCT | GCT | 130 | |
| 21 | Phe | Ala | Ile | Leu | Leu | Ile | Ala | Trp | Leu | Ala | Gly | Phe | Ser | Ser | Arg | Leu | Phe | Ala | Val | 40 | |
| 131 | TTG | CGC | ACC | TCT | GCT | TAAT | CGC | TCT | GCT | GCT | GCG | CGA | TTT | TCT | TCT | CGC | TCT | TCT | CGC | 190 | |
| 41 | Arg | Phe | Glu | Ser | Ile | Ile | His | Glu | Phe | Asp | Pro | Tyr | Phe | Asn | Tyr | Arg | Ala | Thr | Ala | 60 | |

```

421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
1331 GTGGCTTTTCGGACTGTTGGATGTTGTTCTCGAAGAGGATTGCTTAAGCGAATGGGC 1390
441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerLysGluLysThr 460
1391 ACAGCGCATPAGCGCAGCCAGCAGAGTGGATGAGTGAAGTTCATTTGAGAGAGAGACG 1450
461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
1451 CTGTACGCAAGGCTGGCAAGCTGAAGCATCGTACTAAGCATGCTACGATGATGCCAGCAGGATACT 1510
481 GlyValSerSerAsnLeuLysSerLeuValLysLeuAlaValLeuMetLeuLeuMetMet 500
1511 GCGGTCAGCTCAACCTGAGAGATGTTGTTATTTTGGCCGTTCTATGCTGTTGATGATG 1570
501 PheAlaValHisCysThrTyrValThrSerAsnAlaTyrSerProSerLysLeuValLeu 520
1571 TTGCGTGTCCACTGACAGCTGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTTGCTTG 1630
521 AlaPheHisAsnSerGlnAspGlySerArgAsnLeuLeuAspAspPheArgGluAlaTyr 540
1631 GCTTTCCACACAGTCAGATGGATCCCGCAACATTTTAGACGATTTCAAGAGAGGCTTAC 1690
541 TyrTyrLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTyrPheAspTyrGly 560
1691 TACTGGCTTTCCAGACACTGCCGATGATGCTGCGGTTATGTTCTGGTGGGATTACGGA 1750
561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTyrAsnAsn 580
1751 TACCAGATAGCGGGAATGGCAACAGACAGCGGTAGTGGATATAATATAGTGGAACAT 1810
581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
1811 AGTCACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGATGCTACGAAAT 1870
601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
1871 ATGACATCTCTGACGTGGACTACGTTTGTGTGATCTTTGCGGTGTGATCGGCTATCT 1930
621 GlyAspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLys 640
1931 GCGCATGATATCAACAGTTCTCTGTGGATGCTCCGAATTTGCTGAGGAGAGCATCCCAAG 1990
641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
1991 GACATTAAGAAAGCGATTACTTTTACCAGCGCGGTGAATTCAGGGTAGATGCCGAAGT 2050
661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
2051 GCTCCGCGCCCTCCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTG 2110
681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
2111 AGTTTGGACTACAGAGGTCATCTCGATATGATCGACACATAACCGCGCTATTTGGGAAT 2170
701 LysAspPheAspLeuThrTyrLeuGluAlaTyrThrThrGluHisTyrTyrLeuValArg 720
2171 AAGGACTTCGATCTGACCTACCTGAGGAGGCGCTACACACAGAACTGCTGCTTGTTCG 2230
721 IleTyrArgValLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
2231 ATCTATAGGTTGAAGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 2290
741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgTyrLysGlyTyrIle 760
2291 ACGATTCTCCAGCAAACTTCAATTCGAGAAAGAACTCTAAGCGTCCGAAGGGCTACATA 2350
761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
2351 CGAAACCGACCGGTTGTTGTTTAAAGGGAARACCAACCTTGAAA 2392
```

RESULT 2
AX799088
LOCUS
DEFINITION
Sequence 7 from Patent WO03054008.
ACCESSION
AX799088
VERSION
AX799088.1 GI:37605061
KEYWORDS
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
SOURCE
Perreault, C. and McBride, K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 7 03-JUL-2003;
Comptatigene Inc. (CA)
Location/Qualifiers
1..2417
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/notice="SFT3, Acc# AF132552"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2417
Score: 774.00 Matches: 774
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-028-384-8 (1-774) x AX799088 (1-2417)
QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 71 ATGATCGAGCGCGAGAGTGTGAACAGCAGAGGTGGCTGGCTACAGCAGCTTAACACC 130
QY 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 131 TTCGCATCTCTGCTAAATCGGCTGGCTGGCGGGATTTTCTCTCGCTCTTCGGCGCTATC 190
QY 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
Db 191 CGTTTCGATCGATTATCCATGAGTTTGAATCGGTGGTTCAACTACCGGGCCACCGCTTAC 250
QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluAlaTyrTyrPro 80
Db 251 ATGTGCGAAGATGGTTGGTACAACTTCTCAACTGGTTCGACGAGCGCGATGATATCCG 310
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 311 CTCGCGAGATTTGGCGGCTGACCTCTATCCGCGCTGATGATTAGTCCGCGGGAATC 370
QY 101 HisTyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 371 CATTTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTCGCTGTTCTG 430
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
Db 431 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTTACCTGCTGACCAAGAGGCTGG 490
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 491 TCCGCGGCGCGCGGCTCTTCGCGCGCAGCTTTCATCGCCATCGCTCGCTGCTACATCAGT 550
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 551 AGTTCGTTGGCTGGATCGTACGATPACAGAGGCAATTCGCATATTCGCCCTCGCATTCACC 610
QY 181 TyrPheLeuTyrValArgSerValLysThrGlySerValPheTyrSerAlaAlaAla 200

| | | | |
|------|----|--|------|
| 611 | Db | TACITCTCTGGTGGGTGGCTCAGTGAAGACTGGATCCGGTGTCTGGTGGCGCGACGGCT | 670 |
| 201 | Qy | LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleAsnLeuLe | 220 |
| 671 | Db | TTGTGCTCTACTTCTACATGGTGTCCGCCTGGGGTGGCTACGTGTTCATCATCAACCTGATA | 730 |
| 221 | Qy | ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer | 240 |
| 731 | Db | CCCCCTGCACGCTTTCGTACTGCTCATTTATGGCAGGTACTCGCCCGGCTCTCTGACCAGC | 790 |
| 241 | Qy | TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe | 260 |
| 791 | Db | TACAGCACCTTCTACATCTCTGGCATGTGTGTCTCCATGCAGATCCCTCTCGTGGGATTC | 850 |
| 261 | Qy | GlnProIleArgThrSerGluHisMetAlaLeuGlyValPheValLeuLeuMetAla | 280 |
| 851 | Db | CAACCGATACGCACACAGTGAACACATGGCTGGCTGGGAGTGTGTGTCTCTTATGGCC | 910 |
| 281 | Qy | ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgIlysLeuPhe | 300 |
| 911 | Db | GTGGCCACCTTGGCCCAATTGTCAGTCCGTCGTGCGCACAGAGTTCGGGAAGCTGTTC | 970 |
| 301 | Qy | IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuLeuThrMet | 320 |
| 971 | Db | ATCGTGGCGGGATTGCTGTGGCGGCTTGGCGCTTTTGTGGCCGCTGCTGGTGCTCACCATG | 1030 |
| 321 | Qy | LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla | 340 |
| 1031 | Db | CTGGGCGTTGTGGCCCGTGGAGTGGACGCTTCTACTCGCTGGGGATCTGGCTACGCC | 1090 |
| 341 | Qy | LysIleHisIleProIleIleAlaSerGluHisGlnProThrThrTrpPheSer | 360 |
| 1091 | Db | AAGATCCCAATTCCCATCATGTCATCCGTTGCGAGCATCACGCCACCATCTGGTCTCG | 1150 |
| 361 | Qy | PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle | 380 |
| 1151 | Db | TTCTTCTTTGATCTGCACATCCTGGTGTGGCGCTTCCCAGTGGGAGTGGTACTGCATC | 1210 |
| 381 | Qy | LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe | 400 |
| 1211 | Db | AACGACATCAACACAGCGCGTTCCTGGTGTCTGACGCATCAGTCCGGTTCCTTCTTC | 1270 |
| 401 | Qy | AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly | 420 |
| 1271 | Db | GCTGGTGTGATGTGGCTTGAATTTGACCCCTCACCGCGTGTGTGTCATGTGGCGGA | 1330 |
| 421 | Qy | ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerIysArgMetGly | 440 |
| 1331 | Db | GTGGCCCTTTTCGGGACTTTGGATGTGTTCCTCAAGAGGATTCGTCTAAGCGAATGGGC | 1390 |
| 441 | Qy | ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluIysIysThr | 460 |
| 1391 | Db | ACAGCCATAAGCGACGCCACCGAAGTGGATGAAGCTTGAGGATTCATTGAGAAGAAGCG | 1450 |
| 461 | Qy | LeuTyrAspLysAlaGlyLysLeuLysHisArgThrIlyHisAspAlaGlnGlnAspThr | 480 |
| 1451 | Db | CTGTACCAAGGCTGGCAAGCTGAAGCATCGTACTAAGCATCATGCCACGACGGATACT | 1510 |
| 481 | Qy | GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet | 500 |
| 1511 | Db | GGCGTCAGCTCCCAACCTCAAGAGTATTGTATTATTTGGCCGCTCTAATGCTGTGTGATG | 1570 |
| 501 | Qy | PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu | 520 |
| 1571 | Db | TTGCTGTCCACTGCACGTGGGTGACCAAGCAATGCCCTACTCCAGTCCCTCCATGTCTTG | 1630 |
| 521 | Qy | AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheAaGGluAlaTyr | 540 |
| 1631 | Db | GCTTTCACAAACAGTCAAGATGGATCCCGCAACATTTTAGACGATTTACAGAGGGCTTAC | 1690 |
| 541 | Qy | TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly | 560 |

| | | | |
|-------------------|--|--|-----------------|
| 1691 | TACTGGCTTTCCGAGAACAC | TGCGCGATGATGCTCGCGTATATGTTCTTGGTGGATTACGGA | 1750 |
| 561 | TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn | | 580 |
| 1751 | TACCAGATAGCGGGAATGGCAAA | CAGAACGACGCTAGTGGATAATAATACGTGGAAACAAT | 1810 |
| 581 | SerHisIleAlaLeuValGlyLeuAlaMetSerSerThrGluGluLysSerTyrGluIle | | 600 |
| 1811 | AGTCACATAGCGCTGGTGGCAAGCAAGTCTTCA | CCGAGGAGAGACTCTTACGAAATT | 1870 |
| 601 | MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer | | 620 |
| 1871 | ATGACATCTCTTACGTGACGTACGTTT | TGGTGTATCTTTGGCGGTGTCATCGGTATTCT | 1930 |
| 621 | GlyAspAspIleAsnLysPheLeuTrpMetValargIleAlaGluGlyGluHisProLys | | 640 |
| 1931 | GGCGATGATATCAACAAGTTCTCTGGGATGGTCGAAT | TGCTGAGGGAGAGCATCCCAAG | 1990 |
| 641 | AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheAsgValAspAlaGluGly | | 660 |
| 1991 | GACATTAGGAAGCGATTACTTTTACCGACCGCGTGAATT | CAGGCTAGATGCCAAGGT | 2050 |
| 661 | AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu | | 680 |
| 2051 | GCCTCGCGCCCTGCTCAACTGCGCTTATGACAAATTAAGCTACT | CAGAAATTCGGGAATTG | 2110 |
| 681 | LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn | | 700 |
| 2111 | AACTTGGACTACAGAGGTCCTATCGATATGATCGACACGCTAACCGCGTCAAT | TGGGAAT | 2170 |
| 701 | LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValarg | | 720 |
| 2171 | AAAGGACTTCGATCTGACCTACCTGGAGGAGGCGCTACACACAGAAACAT | CTGGCTTTGTCG | 2230 |
| 721 | IleTyrArgValLysIleProHisGluPheAsnArgProSerLeuLysThrLysGluArg | | 740 |
| 2231 | ATCTATAGGTGAGAGCGCGCATGAGTTCATATAGCATCATCACTGAAGACCAAGGAGAGA | | 2290 |
| 741 | ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle | | 760 |
| 2291 | ACGATTCTCCAGCAAACTTCATTT | CGAGAAAGAACTCTTAAGCTCGCAAGGCTACATA | 2350 |
| 761 | ArgAsnArgProValValValLysGlyLysArgThrLeuLys | | 774 |
| 2351 | CGAAACCGACCGGTTGTTGTTTAAAGGAAACCAACCTTGAAA | | 2392 |
| RESULT 3 | | | |
| CO589353 | | 2699 bp | DNA |
| LOCUS | Sequence 17111 from Patent WO0171042. | | linear |
| DEFINITION | | | PAT 02-FEB-2004 |
| ACCESSION | CO589353 | | |
| VERSION | CO589353.1 | GI:41648215 | |
| KEYWORDS | | | |
| SOURCE | Drosophila sp. | | |
| ORGANISM | Drosophila sp. | | |
| | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | |
| | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| | Ephydroidea; Drosophilidae; Drosophila. | | |
| REFERENCE | 1 | | |
| AUTHORS | Venter, J.C., Adams, M., Li, P.W. and Myers, B.W. | | |
| TITLE | Detection kits, such as nucleic acid arrays, for detecting the | | |
| | expression of 10,000 or more Drosophila genes and uses thereof | | |
| JOURNAL | Patent: WO 0171042-A 17111 27-SEP-2001; | | |
| | PE Corporation (NY) (US) | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..2699 | | |
| | /organism="Drosophila sp." | | |
| | /mol_type="unassigned DNA" | | |
| | /db_xref="taxon:7242" | | |
| ORIGIN | | | |
| Alignment Scores: | 0 | Length: | 2699 |
| Fred. No.: | | | |

Db 2120 AAGTTGGACTACAGAGGTCCATCTGGATATGATCGCACACGTAACGCCGCTCATTTGGGAAT 2179
 QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrGluHisTrpLeuValArg 720
 Db 2180 AAGACTTCGATCTGACTACCTCTGGAGAGGCTTACACACAGAACACTGGCTTGTTCGC 2239
 QY 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 Db 2240 ATCTATAGGTGAAGAAGCGCATGAGTTCAATAGACCATCACTGAAGACCAAGAGAGA 2299
 QY 741 ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIle 760
 Db 2300 ACGATTCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTCGCAAGGGCTACATA 2359
 QY 761 ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774
 Db 2360 CGAAACCGACCGGTGTGTGTTAAGGGAACGAACCTTGAAA 2401
 RESULT 4
 CQ589352/c
 LOCUS CQ589352 4922 bp DNA linear PAT 02-FEB-2004
 DEFINITION Sequence 17110 from Patent WO0171042.
 ACCESSION CQ589352
 VERSION CQ589352.1 GI:41648214
 KEYWORDS Drosophila sp.
 SOURCE Drosophila sp.
 ORGANISM Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1
 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 TITLE Detection kits, such as nucleic acid arrays, for detecting the
 AUTHORS expression of 10,000 or more Drosophila genes and uses thereof
 JOURNAL Patent: WO 0171042-A 17110 27-SEP-2001;
 PE Corporation (NY) (US)
 FEATURES
 source
 1..4922
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 4922
 Score: 464.00 Matches: 464
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.95% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-8 (1-774) x CQ589352 (1-4922)
 QY 1 MethAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 Db 3843 ATGAATCGGACGCCGAAGATCTCTCAACAGCAAGTGGCTGGCTACAGACGCTTAATCACC 3784
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 Db 3783 TTCCCATCTCTTAATCGGCTGGCTGGCCGGAATTTCTCTCGCCTCTTCCCGCTCATC 3724
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
 Db 3723 CGTTTCAGTGCATATATCCATGATGTTGATCCCGTGTTCAACTACCGGGCCACCGCCTAC 3664
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
 Db 3663 ATGGTCAGAGATAGTGTGGTACACTTCTCACTGGTTTCGACGAGCGCGCATGGATCCG 3604
 QY 81 LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 3603 CTCGGCAGGATTTGGCGGCTACCGTCTATCCCGGCTGATGATTACGTCCGGCGGAATC 3544

QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 3543 CATTGGCTGCTGCACGCTACTCAACATACGGTCCATATTCGTGACATCTCGGTGTTCTG 3484
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 3483 GGGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGCTGG 3424
 QY 141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 3423 TCGCGGGGGCGGGCTCTTCGCCGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 3364
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 3363 AGGTCCGTGGCTGGATCGTACGATAACGAGGGGCAATGCCATATTCGCCCTCGCATTCACC 3304
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla 200
 Db 3303 TACTTCTGTGGTGGCTCAGTGAAGACTGGATCCGTTCTGGTGGCGCGACCGCT 3244
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220
 Db 3243 TTGTCTCTACTTCTACATGGTGTCCGCTGGGTGGCTACGTGTTCATCATCAACCTGATA 3184
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 Db 3183 CCCCTGCACGCTTTCGTACTGCTCATTTAGGCGAGTACTCGCGCGCTGTGTGACCCAGC 3124
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 Db 3123 TACAGCACCTTCTACATCTGGACTGCTGTCTTCATGAGATCCCTCTCGTGGGATTC 3064
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 Db 3063 CAACCGATACGACCGACGATGACATGCTGCTGGTGGAGGTGTGTGCTCTCTTATGGCC 3004
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 Db 3003 GTGGCCACCTTCGGCCATTTGCAGTCCGTGCTGCGCGCAACAGAGTTCGGAGAGTGTTC 2944
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 Db 2943 ATCGTCGGCGGATTCGTGGTGGCGCTGGCGCTCTTTGTGGCGCTGTGTGCTCACCANG 2884
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 Db 2883 CTGGCGGTGTGGCCCCCGGTGGAGTGGAGCTTCTACTCGCTGGGATACTGGGTACGCC 2824
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 Db 2823 AAGATCCACATTCCTCCATTCATTCGATCCGTTGCGAGCATCAGCCACCATTTGGTTCG 2764
 QY 361 PhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 Db 2763 TTCTCTTTTGATCTGCACATCTCTGTGTGGCTTCCAGTGGGAGTGTGTGTACTGTGATC 2704
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 Db 2703 AAGCAGATCAACGACGAGCGGGTTCGTGGTGTGTGTACGCCATCAGTGGGTACTTTC 2644
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 2643 GCTGGTGTGATGGTGGCTTGTATGTTGACCTCAGCGCGTGGTGTGTGATGTGGCGGA 2584
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 Db 2583 GTGGCTTTTCGGGACTGTGGATGTGTCTCTGCAAGAGGATTCGTCTAAGCGAATGGGC 2524
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 Db 2523 ACAGCCATTAAGCGCAGCCACCGAAGTGGATGAGCTGAGGATTCATTGAGAAGAAGACG 2464
 QY 461 LeuTyrAspLys 464


```

Db      2463 CTTGACGACAG 2452
|||||
RESULT 5
AC007853/c
LOCUS
DEFINITION
AC007853
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 162921)
REFERENCE
AUTHORS
Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banazon, J., Beeson, K.Y., Busan, D.A.,
Carlson, J.W., Center, A., Chame, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Ferreira, S., Frisbe, J., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, C., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Scheeler, F.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 3R, region 96B-96C
Unpublished
2 (bases 1 to 162921)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomctan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacieb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 27, 2001 this sequence version replaced gi:5670481.
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
FEATURES
Source
1. 162921
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="96B-96C"
/clone="BACR03L02 (D766)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 59.95% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-8 (1-774) x AC007853 (1-162921)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 108481 ATCAATCGGACGCGGAGATGCTGAACAGCAGAGTGGCTGGCTACAGCAGCTTAATCACC 108422
Qy 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 108421 TTCCCATCTGCTAATCGCTGGCTGGCGGAGATTTCCTCTCGCTCTTCGCGGTCATC 108362
Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
Db 108361 CGTTTCAGTTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGGGCCACCCCTAC 108302
Qy 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db 108301 ATGGTCAGATGGTTGGTACAACTTCCTCAACTGTTTCGACGAGCGCATGGTATCG 108242
Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 108241 CTCGGCAGGATTGTGGCGGATACCGCTATATCCCGGCTGATGATTACGTCGCGCGGAATC 108182
Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 108181 CATTGGCTGCTGCACGTACTCAACATACCGGTCATATTCGTGACATCTCGGTGTCTCTG 108122
Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGluLeuTrp 140
Db 108121 GCGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTACCTACCAAGAGAGCTGTGG 108062
Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 108061 TCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCATCTGCTGGCTGGCTACATCAGT 108002
Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 108001 AGGTCGGTGGTGGATCGTAGATAACGAGGCGATTGGCATATTCCGCTCAGTTCACC 107942
Qy 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAla 200
Db 107941 TACTTCTGTGGTGGCTGCTCAGTAGACTCGGCTGCTGCTGGCGCGAGCGCT 107882
Qy 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
Db 107881 TTGTCTACTTCTACATGCTGCTGGCTGGCTGGCTAGCGTTTCATCATCACTGATA 107822
Qy 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
Db 107821 CCCCTGACGCTTCGTACTCTATTATGGCGAGGTACTCGCGCGCTCTCTGACGAGC 107762
Qy 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
Db 107761 TACAGCAGCTTCTACATCTCGGACTGCTGTTCTCCATGAGATCCCTCTCGTGGGATTC 107702
Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuMetAla 280
Db 107701 CAACCGATACGACACGAGACACATGCTGGCTGGGAGTGTGTGTGCTCTCTTATGGCC 107642
Qy 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
Db 107641 GTGGCCACCTTGCCTGATTTGCGAGTCCGTGCTGTCGCGACGAGTTCGGAAGCTGTTTC 107582
Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
Db 107581 ATCGTCGCGGATTCGCTGGTGGCGCTTGGCGCTCTTGTGGCCGCTCGTGGTGTCCACCATG 107522
Qy 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340

```

Alignment Scores:
Pred. No.: 0
Score: 464.00
Length: 162921
Matches: 464

Db 107521 CTGGGCGTTGGCCCGTGGAGTGGACGCTTCTACTCGCTGGGATGACTGCTACGCC 107462

QY 341 LysleHisIleProIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360

Db 107461 AAGATCCACATTCCTCATTCGTCGTCGAGCATCAGCCACCACTTGGTCTCG 107402

QY 361 PhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTrpCysIle 380

Db 107401 TTCCTTTGATCTGCACATCTCTGGTGGCTTCCAGTGGAGTGGTACTGCATC 107342

QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTrpPhe 400

Db 107341 AAGCAGATCAACAGCAGCGCGTTTCTGCTGCTGTACGCCATCAGTCCGTTTACTTC 107282

QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420

Db 107281 GCTGGTGTGATGTGGCTTTGATGTTGACCTCAGCCGGTGGTGGCATGTGGCCGGA 107222

QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerIysArgMetGly 440

Db 107221 GTGGCTTTTCGGACACTGTTGGATGTTCTCTCAAGAGGATTCGTCTAAGCGAATGGGC 107162

QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluIysIysThr 460

Db 107161 ACACCATAGCCAGCAGCCACCGAGTGGATGAGTTCAGATTCATTCAGAGAGAGAG 107102

QY 461 LeuTyrAspIys 464

Db 107101 CTGTACGACAG 107090

RESULT 6

AC008206/c 181132 bp DNA linear INV 24-FEB-2001

DEFINITION Drosophila melanogaster, chromosome 3R, region 96B-96B, BAC clone

ACCESSION BAC003115, complete sequence.

VERSION AC008206

KEYWORDS HTG.

SOURCE AC008206.10 GI:13122717

ORGANISM Drosophila melanogaster (fruit fly)

REFERENCE 1 (bases 1 to 181132)

AUTHORS Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Buam,D.A., Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreanek,D., Farfan,D., Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwan,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puti,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svirska,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

TITLE Sequencing of Drosophila chromosome 3R, region 96B-96B

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 181132)

AUTHORS Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazef,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomoran,M.A., Marda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,B., Sequera,A., Sethi,H., Snir,F., Svirska,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

TITLE Direct Submission

JOURNAL Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT On Feb 24, 2001 this sequence version replaced gi:7208834.

Sequence submitted by:

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgs@fruitfly.berkeley.edu.

FEATURES

source

1..181132
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="96B-96B"
/clone="BACR03115 (D765)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"

ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 181132
Score: 464.00 Matches: 464
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 59.95% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-8 (1-774) x AC008206 (1-181132)

QY 1 MetAsnArgThrProIysMetLeuAsnSerIysValAlaGlyTyrSerSerLeuIleThr 20

Db 171813 ATGAATCGGACGCCGAGATGCTGAACGAAGTGGTGGCTACAGCAGCTAATCACC 171754

QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40

Db 171753 TTCCGCATCTCTGTAATGCTTGGCTGGCCGGATTTTCCTTCGCCCTCTTCGGGTCATC 171694

QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60

Db 171693 CGTTTCGAGTCGATATTCATGAGTTTGCATCGTGGTTCACACTACCGGCGCCGCTAC 171634

QY 61 MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80

Db 171633 ATGGTGCAGAAATGTTGTACAACTTCTCACTGGTTCGACGAGCGCGATCGTATCCG 171574

QY 81 LeuGlyArgIleValGlyGlyThrValTrpProGlyLeuMetIleThrSerGlyGlyIle 100

Db 171573 CTCGGCAGGATTTGGGGGGTACCGCTCTATCCCGGCTGTAGTATACGTCGGGGAATC 171514

QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120

Db 171513 CATTCGCTGCTCAGTACTCAACATACCGTTCATATTCGTGACATCTCGGTGTCTCG 171454

QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140

Db 171453 GCGCCGATTTTCAGTGGCTGACCTCCACTCTCCACTACCTACCTGTCGACCAAGAGCTGTGG 171394

QY 141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

Db 171393 TCGGCGGGCGCGGCTCTTTCGGCGCCAGCTTCTATCGCCATCGTGGCTGCTACATCAGT 171334

QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180

Db 171333 AGTCCGCTGGCTGGATCGTACGATAACGAGGCGATTCGCATATTCGCCCTCGAGTTCCACC 171274

QY 181 TyrPheLeuTrpValArgSerValIysThrGlySerValPheTrpSerAlaAlaAla 200

Db 171273 TACTTCTGTGGTGGCTCAGTGAAGACTGATCGGTGTTCTGTGCGCGCAGCGCGCT 171214


```
Db 130700 CATTGGCTCTCCAGTACTCAACATACCGGTCCATATTCGTGACATCTGGTGTCTCGT 130641
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
Db 130640 GGGCCGATCTTCAGTGGCTGAGCTCCATCTCCACCTACTCTGTCACCAAGAGCTGTGG 130581
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheAlaIleValProGlyTyrIleSer 160
Db 130580 TCCGGGGGGCGGCGCTCTTGGCCGCCAGCTTCATCGCCATCGTGGCTGTACATCAGT 130521
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 130520 AGTCCGTGGCTGATCGTACGATACAGAGGCGATTCGCCATTCGCCCTCCAGTTCACC 130461
QY 181 TyrPheLeuTyrValArgSerValIleThrGlySerValPheIlePheAlaAlaAla 200
Db 130460 TACTTCCTGGTGGGTCGCTAGTGAAGACTGATCCGTGTCTGGTGGCGCAGCGCT 130401
QY 201 LeuSerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIle 220
Db 130400 TTGTCTTACTTCTACATGGTGTCCGCTGGGGTGGCTACGGTTCATCATCAACCTGATA 130341
QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
Db 130340 CCCCCTGCAGCTTCTGACTCTCATATGGCCAGGTACTCGCCGCGTCTGCTGACCAGC 130281
QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
Db 130280 TACAGCACCTTCTACATCTGGGACTGCTGTCTCCATGAGAGTCCCTTCGTGGGATTC 130221
QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
Db 130220 CAACCGATACGACAGTGAACATCGCTGGCTGGAGTGTGTGCTCTTATGGCC 130161
QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
Db 130160 GTGCCACCTTCGCCCATCTTGCAGTCCGCTGCTGCGGCAACGAGTTCGGGAAGCTGTT 130101
QY 301 IleValGlyGlyLeuLeuValGlyValPheValAlaValValLeuLeuThrMet 320
Db 130100 ATCGTGGCGGATGCTGGTGGGGTGGCGTCTTGTGGCCGCTGGTGGCTCACCATG 130041
QY 321 LeuGlyValValAlaProTyrSerGlyArgPheTyrSerLeuTyrPheSerThrGlyTyrAla 340
Db 130040 CTGGCGCTTGTGGCCGCTGGAGTGGAGCTTCTACTCGCTGGGATACCTGGCTACGCC 129981
QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrPheSer 360
Db 129980 AAGATCCACATTCACATCATGTATCGTGGAGCATCAGCCCACTTGGTGTCTCG 129921
QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrTyrCysIle 380
Db 129920 TTCTTCTTGTATCTGCACATCTGCTGGCTTCCAGTGGAGTGTGGTACTGCATC 129861
QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
Db 129860 AACAGATCAACAGCAGCGCGTTCCTGGTGGTCTGTACGCCATCAGTGGGTTACTTC 129801
QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
Db 129800 GCTGGTGTATGTGGTGTGTGTGTGACCTCACCGCGGTGGTGTGTGTGTGGCGGA 129741
QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
Db 129740 GTGGCTTTTCGGGACTGTGGATGTTCCTGGTGGTGTGTGTGTGTGTGTGTGTGTGTGT 129681
QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460
Db 129680 ACAGCCATAAGCCACCCAGAGTGGATGAGTGAAGTTCAGGATTCATTGAGAAGAGACG 129621
QY 461 LeuTyrAspLys 464
Db 129620 CTGTACGACAAG 129609
```

RESULT 8
AC018145/c
LOCUS AC018145 75650 bp DNA linear HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
AC018145
ACCESSION AC018145
VERSION AC018145.1 GI:6553046
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 75650)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10213542 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..75650
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 75650
Score: 390.00 Matches: 463
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 1
Query Match: 50.39% Indels: 1
DB: 2 Gaps: 0
US-10-028-384-8 (1-774) x AC018145 (1-75650)
QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 21487 ATGATCGAGCGCGGAGATCTGACAGCAGAGTGGCTGGCTACAGCAGCTATATCACC 21428
QY 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 21427 TTCCGATCTCTGCTAAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGCTCATC 21368
QY 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
Db 21367 GCTTTCGATCGATATTCATGAGTTGATCCGTTGGTTCAACTACCGGGCCAGCGCTAC 21308
QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80
Db 21307 ATGCTGCAGATGTTGGTACAACTCTCTCACTGGTTCGACGAGCGGATGTATCCG 21248
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 21247 CTCGGCAGGATGTGGGCGGTACCGTCTATCCCGCGCTGTATGATTCGCGCGGAATC 21188
QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 21187 CATTTGGCTGTGACGTACTCAACATACCGGTCCATATTCGTGACATCTGGGTGTTCTG 21128
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
Db 21127 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGTGTGTGTGTGTGTGTGTGT 21068
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 21067 TCCGGCGGCGCGGCTCTTCGCGCGCAGCTTTCATCGCCATCGTGGCTGTGCTACATCAGT 21008


```

DEFINITION      Sequence 24553 from Patent WO0171042.
ACCESSION       CQ596795
VERSION         CQ596795.1  GI:41653176
SOURCE          Drosophila sp.
ORGANISM        Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                Ephydroidea; Drosophilidae; Drosophila.
REFERENCE       1
AUTHORS         Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
TITLE           Detection kits, such as nucleic acid arrays, for detecting the
                expression of 10,000 or more Drosophila genes and uses thereof
                Patent: WO 0171042-A 24553 27-SEP-2001;
                PE Corporation (NY) (US)
FEATURES        location/Qualifiers
                1..2785
                /organism="Drosophila sp."
                /mol_type="unassigned DNA"
                /db_xref="taxon:7242"
ORIGIN
Alignment Scores:
Pred. No.:      1.21e-75      Length:      2785
Score:          83.00         Matches:      83
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    10.72%       Indels:        0
DB:             6            Gaps:          0

US-10-028-384-8 (1-774) x CQ596795 (1-2785)
Qy      603 SerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 622
Db      2784 TCCTTGACGTGGACTACGTTTGCTGATCTTTGGCGGTGTGATCGGCTATTCTGGCAT 2725
Qy      623 AspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIle 642
Db      2724 GATATCAACAAGTTCCTGTGGATGTCGGAATTCGTGAGGAGCATCCCAAGGACATT 2665
Qy      643 LysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaPro 662
Db      2664 AAGGAAGAGCATTAATTCTACGACCCGCGTGAATTCAGGTTAGATCCGAGGTCTCCG 2605
Qy      663 AlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeu 682
Db      2604 GCCCTGCTCAACTGCCCTTATGTACAAATTAAAGTACTACAGATTCGGGATTCAGATTG 2545
Qy      683 AspTyrArg 685
Db      2544 GACTACAGG 2536

RESULT 11
UI3019
LOCUS          UI3019          54118 bp      DNA      linear      INV 13-JAN-2003
DEFINITION     Caenorhabditis elegans cosmid T12A2, complete sequence.
ACCESSION      UI3019
VERSION        UI3019.1  GI:912482
KEYWORDS       HTG.
SOURCE         Caenorhabditis elegans
ORGANISM       Caenorhabditis elegans
               Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
               Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE       1 (bases 1 to 54118)
AUTHORS         Waterston,R.
TITLE           Genome sequence of the nematode C. elegans: a platform for
               investigating biology. The C. elegans Sequencing Consortium
               Science 282 (5396), 2012-2018 (1998)
JOURNAL        99069613
MEDLINE        9851916
PUBMED         9851916
REFERENCE       2 (bases 1 to 54118)
AUTHORS         Latreille,P.
TITLE           The sequence of C. elegans cosmid T12A2

```

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
3 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (06-MAR-1996) Genome Sequencing Center, Washington
University

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
4 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (26-JUN-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
5 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (09-AUG-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
6 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
7 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (28-AUG-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
8 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (21-NOV-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

JOURNAL REFERENCE AUTHORS TITLE

Unpublished (2001)
9 (bases 1 to 54118)
Waterston,R.
Direct Submission
Submitted (11-JAN-2003) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA

JOURNAL REFERENCE AUTHORS TITLE

On Jul 27, 1995 this sequence version replaced gi:529354.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one mi3 subclone.

For a graphical representation of this cosmid sequence and its
analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T12A2;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C23G10, 200 bp overlap; the 3' cosmid is C18P10,
15000 bp overlap. Actual start of this cosmid is at base position 1
of T12A2; actual end is at 39122 of T12A2.

NOTES:

QY 551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

us-10-028-384-8.Oligo.rge

Thu Dec 16 16:25:14 2004

DEFINITION Sequence 402 from Patent EP1067182.
 ACCESSION AX136480
 VERSION AX136480.1 GI:14272884
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.
 TITLE Secretory protein or membrane protein
 JOURNAL Patent: EP 1067182-A 402 10-JAN-2001;
 Helix Research Institute (JP)

FEATURES
 source
 1..764
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:
 Pred. No.: 2.44e-33 Length: 764
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x AX136480 (1-764)

| | | | |
|----|-----|--|-----|
| Qy | 551 | AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr | 570 |
| Db | 268 | GCACGAGTAATGTCITGGTGGGATTATGGCTACAGATAGCTGGAAATGGCTAATAGAACT | 327 |
| Qy | 571 | ThrLeuValAspAsnAsnThrTrpAsnSerHisIleAlaLeuValGlyLysAlaMet | 590 |
| Db | 328 | ACGTTGGTGGATTAATACACCTGGATACAGCCACATAGCACTGGTGGGAAAAGCTATG | 387 |
| Qy | 591 | SerSer | 592 |
| Db | 388 | TCTTCT | 393 |

Search completed: December 14, 2004, 22:47:49
 Job time : 8937.94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 891.777 Seconds
(without alignments)
4556.130 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRPEKMLNSKAVGVSSLI.....RRKGYIRNPVVKGKTLK 774

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10028384/runat.14122004.131645.6790/app_query.fasta.1.3740
-DB=N Geneseq 23Sep04 -QPMT=fastap -SUFFIX=Oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10028384 @CEN 1.1.1986 @runat.14122004.131645.6790 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq 23Sep04:
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 774 | 100.0 | 2417 | 10 | ADD94789 |
| 2 | 774 | 100.0 | 2699 | 4 | ABLI3247 Drosophila |
| C 3 | 464 | 59.9 | 4922 | 4 | ABLI3246 Drosophila |
| C 4 | 208 | 26.9 | 2953 | 4 | ABLI18224 Drosophila |
| C 5 | 83 | 10.7 | 2785 | 4 | ABLI18208 Drosophila |
| 6 | 42 | 5.4 | 433 | 6 | ABLI37128 Human col |

ALIGNMENTS

RESULT 1
ADD94789
ID ADD94789 standard; DNA; 2417 BP.
XX
AC ADD94789;
XX
DT 29-JAN-2004 (first entry)
XX
DE Drosophila melanogaster STT3 gene sequence.
XX
KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide;
KW fruit fly; Gene; ds; STT3.
XX
OS Drosophila melanogaster.
XX
PN WO2003054008-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-CA001967.
XX
PR 20-DEC-2001; 2001US-00028384.
XX
PA (COMP-) COMPATIGENE INC.

| | | | | | | |
|----|----|-----|------|----|----------|---------------------|
| 7 | 42 | 5.4 | 483 | 3 | AAA43486 | Aaa43486 Mouse sec |
| 8 | 42 | 5.4 | 507 | 3 | AAC01967 | Aac01967 Human sec |
| 9 | 42 | 5.4 | 754 | 5 | AAF93958 | Aaf93958 Primer bp |
| 10 | 42 | 5.4 | 787 | 4 | AAR07526 | Aar07526 Human cDN |
| 11 | 42 | 5.4 | 2284 | 4 | AAH18021 | Aah18021 Human cDN |
| 12 | 42 | 5.4 | 2481 | 10 | ADD94783 | Add94783 Human SIM |
| 13 | 42 | 5.4 | 2510 | 5 | AAF93772 | Aaf93772 Human cDN |
| 14 | 42 | 5.4 | 2537 | 4 | AAD08289 | Aad08289 Human sec |
| 15 | 42 | 5.4 | 2546 | 2 | AAV44866 | Aav44866 Clone C75 |
| 16 | 42 | 5.4 | 2546 | 5 | AAF98483 | Aaf98483 Human cDN |
| 17 | 42 | 5.4 | 2547 | 4 | AAD08315 | Aad08315 Human sec |
| 18 | 42 | 5.4 | 2660 | 6 | ABQ54750 | Abq54750 Human ova |
| 19 | 42 | 5.4 | 2710 | 10 | ADD94785 | Add94785 Mouse SIM |
| 20 | 41 | 5.3 | 349 | 2 | AAV89737 | Aav89737 EST clone |
| 21 | 34 | 4.4 | 313 | 8 | ABZ20328 | Abz20328 Group III |
| 22 | 34 | 4.4 | 387 | 4 | ABA08397 | Aba08397 Human sec |
| 23 | 34 | 4.4 | 558 | 5 | AAF93612 | Aaf93612 Umbilical |
| 24 | 34 | 4.4 | 1209 | 4 | AAH33264 | Aah33264 Human col |
| 25 | 34 | 4.4 | 1209 | 6 | ABL89850 | Abi89850 Human pol |
| 26 | 34 | 4.4 | 1543 | 2 | AAK85055 | Aak85055 Human sec |
| 27 | 34 | 4.4 | 1543 | 8 | ACD18981 | Acd18981 Novel hum |
| 28 | 34 | 4.4 | 1543 | 12 | ADG78372 | Adg78372 Human sec |
| 29 | 34 | 4.4 | 1543 | 12 | ADN60663 | Adn60663 Human sec |
| 30 | 34 | 4.4 | 1664 | 4 | AAK94164 | Aak94164 Human ful |
| 31 | 34 | 4.4 | 1664 | 12 | ADL30661 | Adl30661 Full leng |
| 32 | 32 | 4.1 | 1848 | 8 | ABT18969 | Abt18969 Aspergill |
| 33 | 32 | 4.1 | 1969 | 8 | ABT18375 | Abt18375 Aspergill |
| 34 | 32 | 4.1 | 2232 | 8 | ABT20789 | Abt20789 Aspergill |
| 35 | 32 | 4.1 | 2603 | 8 | ABT20191 | Abt20191 Aspergill |
| 36 | 32 | 4.1 | 3969 | 8 | ABT17781 | Abt17781 Aspergill |
| 37 | 32 | 4.1 | 4603 | 8 | ABT19595 | Abt19595 Aspergill |
| 38 | 30 | 3.9 | 1114 | 4 | AAH99794 | Aah99794 Human pro |
| 39 | 28 | 3.6 | 281 | 3 | AAA45221 | Aaa45221 Human sec |
| 40 | 28 | 3.6 | 500 | 6 | ABV88332 | Abv88332 Human col |
| 41 | 28 | 3.6 | 616 | 3 | AAZ80248 | Aaz80248 Human col |
| 42 | 27 | 3.5 | 474 | 5 | AAF93633 | Aaf93633 Umbilical |
| 43 | 23 | 3.0 | 388 | 5 | AAF66328 | Aaf66328 Novel hum |
| 44 | 22 | 2.8 | 2855 | 4 | ABL02795 | Abi02795 Drosophila |
| 45 | 22 | 2.8 | 6153 | 4 | ABL02794 | Abi02794 Drosophila |

XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.
 XX DR P-PSDB; ADD94790.
 XX
 PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Claim 6; SEQ ID NO 7; 66pp; English.
 XX
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the Drosophila melanogaster SPT3 gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 2417
 Score: 774.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-10-028-384-8 (1-774) x ADD94789 (1-2417)
 Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
 Db 71 ATGAATCGGACGGCGAAGATGCTGAACAGCAAGTGGCTGGCTACAGACCTATACACC 130
 Qy 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValle 40
 Db 131 TTCGCCATCTGCTAATCGCTGGCTGGCGCGGATTTCTCTCGCTCTTCGCGTCATC 190
 Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrAlaTyr 60
 Db 191 CGTTTCAGATCGATATATCCATGATGATCCGTGGTCAACATACCGGGCCACCGCTAC 250
 Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
 Db 251 ATGATGAGATGTTGGTACAACTTCTCAACTGGTTTCGACGAGCGCGATGGTATCGG 310
 Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 311 CTCGGCAGGATTTGGGGGGTACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATC 370
 Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 371 CATTTGGCTGTCGACGATACCAATACCGGCTCATATTCGTGACATCTCGGTGTTCTCG 430
 Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 431 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 490
 Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

Db 491 TCCGGGGGGCGCGGCTCTTTGCCGCCAGCTTCATCGCCATCGTGGCTGGTACATCAGT 550
 Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 551 AGTTCGGTGGCTGGATCGTACGATAACGAGGCGATTCGCATATTCGCCCTCGAGTTACCC 610
 Qy 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheThrSerAlaAlaAlaAla 200
 Db 611 TACTTCTCTGGTGGCTCGCTCAGTGAAGACTGGATCCGTGTTCTGTGCGCGCGAGCCGCT 670
 Qy 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
 Db 671 TTGTCTTACTTCTACATGGTGTTCGGCTGGGCTGGCTAGTGTTCATCATCAACCTGATA 730
 Qy 221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
 Db 731 CCCTCGACGCTCTTCGTACTGCTCATTATGGCAGGTACTCGCCGCGTGTGTGACCCAGC 790
 Qy 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 Db 791 TAGACACCTTCTACATCTCGGACTGCTGTTCTCCATGCAGATCCCTTCGTGGGATTC 850
 Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 Db 851 CAACCGATACGACACCATGTAACATGGCTGGGCTGGGAGTGTGTTGTCTTATGGCC 910
 Qy 281 ValAlaThrLeuArgHisIleGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 Db 911 GTGGCACCCTTGGCCCATTTGCACTGCTGGCTGCTGCGGCAAGAGTTCCGAGAGCTGTC 970
 Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuLeuThrMet 320
 Db 971 ATCGTGGCGGATTTGCTGGTGGCGGTGGCTGCTTTGTGGCGCTCGTGGTGTCCACATG 1030
 Qy 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 Db 1031 CTGGCGCTTGTGGCCCGCTGGAGTGACCGCTTCTACTCGCTGGGATGACTGGCTACGCC 1090
 Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpPheSer 360
 Db 1091 AAGATCCACATTCCTCATTCATCCGTGGGAGCATCAGCCACACCATCTGGTCTCG 1150
 Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 Db 1151 TTCTTCTTGTATCTGCACATCTCTGGTGGCGCTTCCAGTGGAGTGGTACTGCATC 1210
 Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 Db 1211 AAGCAGATCAACGACGAGCGCTTTCGTGCTGTACGCCATCAGTGGCTTACTTC 1270
 Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 1271 GCTGGTGTATGTTGGTGGCTTGTATGTTGATGTTGATGTTGATGTTGATGTTGATGTTG 1330
 Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 Db 1331 GTGGCTTTCGGGACTTGTGGATGTTGTTCTGTCAGAGAGATTCGCTAAGCGAATGGGC 1390
 Qy 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 Db 1391 ACAGCCATAGCGCAGCCACCGAAGTGGATGAGTGGATTCATTCATGAGAGAGACG 1450
 Qy 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
 Db 1451 CTGTACGACAAAGGCTGGCAGCTGAAGCATGCTACTAAGCATGATGCCGACGAGTACT 1510
 Qy 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
 Db 1511 GCGGTGACCTCCACCTGAAGAGATTTGTTATTTGGCCGCTTCTAATGCTGTTGATGATG 1570
 Qy 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520

Db 1571 TTCGCTGCCACTGCACGTGGGTGACAGCAATGCTTACTCCAGTCCCTCCATGCTTG 1630
 QY 521 AlapheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr 540
 Db 1631 GCTTTCCACACAGTCAAGATGGATCCGCAACATTTTAGACGATTTTCAGAGAGGCTTAC 1690
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly 560
 Db 1691 TACTGGCTTCCAGAACACATGCCGATGATGCTCGGTATGCTTGGTGGATACGGA 1750
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 Db 1751 TACCAGATAGCGGGAATGCAACAGACGAGTGTAGTGGATAATAATACGTGGAACAA 1810
 QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 Db 1811 AGTCATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAT 1870
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 Db 1871 ATGACATCTCTTACGCTGACTACGCTTTTGGTGTATCTTTGGCGGTGTGATCGGCTATCT 1930
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 Db 1931 GCGGATGATCAACAGTCTCTGCGATGCTCGAATGCTGAGGAGAGCATCCCAAG 1990
 QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAlaGluGly 660
 Db 1991 GACATTAAGGAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGATGCCAAGGT 2050
 QY 661 AlaProAlaLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 Db 2051 GCTCGCGCTCTCACTGCTTATGACAAATTAAGCTACTACAGATTCCGGGAATG 2110
 QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 Db 2111 AAGTTGGACTACAGAGTCCATCTGATATGATCGCACATGACGCGTCAITGGGAAT 2170
 QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArg 720
 Db 2171 AAGGACTTCGATCTGACCTACTGAGGAGGCGCTACACCAAGAACACTGCTTGTTCGC 2230
 QY 721 IleTyrArgValLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 Db 2231 ATCTATAGGCTGAAGAGCGCATGATTCATTAATAGACCATCACTGAAGACCAAGGAGA 2290
 QY 741 ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760
 Db 2291 AGATTCTCCAGCAACTTCATTTGAGAAAGAACTCTAAGCGTCCAAAGGGGTACATA 2350
 QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 Db 2351 CGAAACCGACCGGTGTGTTAAGGGAACCAACCACTTGAAC 2392

RESULT 2

ID ABL13247 standard; cDNA; 2699 BP.
 AC ABL13247;
 DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 AC ABL13247;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 PN 27-SEP-2001.
 XX

PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW,
 XX P-PSDB; ABB69144.
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB120511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (AB557737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2699
 Score: 774.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x ABL13247 (1-2699)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 Db 80 ATGAATCGAGCGCGAAGATGCTCAACAGCAAGTGGCTGGCTACAGCAGCTTAATCACC 139
 QY 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 Db 140 TTCGCCATCTCTGTAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGCTAATC 199
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 Db 200 CGTTTCGAGTCGATTTATCCATGATTTGATCCGCTGGTTCACCTACCGGCGCACCGCTAC 259
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrPro 80
 Db 260 ATGGTGCAGAAATGGTGGTACAACTTCTCAACTGGTTCGACGAGCGCGATGGTATCCG 319
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 320 CTCGGCAGAGATTGTGGCGGTACCGTCTATCCCGCGCTGATGATGATGATGATGATGAT 379
 QY 101 HisTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 380 CATTGGCTGTCGACGACTCAACATACCGGTCCATATTCGTACATTCGCGTGTCTG 439
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 440 GCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGCTGTGG 499
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 500 TCCGCGGCGCGCGCTCTTCGCGCGAGCTTTCATCGCCATCGTGGCTGCTACATCAGT 559

PR 11-JUL-2000; 2000US-00614150.
 XX (PEXE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EM;
 XX WPI; 2001-656860/75.
 DR F-PSDB; ABB69143.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 34220; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB857737-
 CC AB872072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
 XX

Alignment Scores:
 Pred. No.: 0 Length: 4922
 Score: 464.00 Matches: 464
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.95% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x ABL13246 (1-4922)

QY 1 MetAsnArgThrProIysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 DB 3843 ATGAATCGGACCGCCGAAGTGTGACAGCAAGGTGGGTGGTGTGACGAGCCTAATCACC 3784
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 3783 TTCGCCATCTCTAATCGCTGGCTGGCCGGAATTTCTCTCGCCCTTCGCCGTATC 3724
 QY 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
 DB 3723 CGTTTCGAGTCGATTATCCATGAGTTTGATCCGTGGTTCAACTACCGGGCCACCGCTAC 3664
 QY 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluAtgAlaTrpTyrPro 80
 DB 3663 ATGGTCAGATGGTTGGTACACTTCTCACTGGTTCGACGAGCGGCATGATCCG 3604
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 3603 CTCGGCAGGATTTGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGGAATC 3544
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 3543 CATTGGCTGCTGCACGTACTCAACATACCGTTCATATTCGTGACATCTGGGTGCTG 3484
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 DB 3483 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGCTGGG 3424
 QY 141 SerAlaGlyValGlyLeuPheAlaIleAsnPheIleAlaIleValProGlyTyrIleSer 160
 DB 3423 TCCGCGGCGCCGCGCTCTTCGCGCGCAGCTTCATCGCCATCGCTGGCTGGTACATCAGT 3364
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 3363 AGTTCGGTGGCTGGATCGTACGATACGAGGGCATTTGCCATATTCGCCCTCAGTTACCC 3304

QY 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAlaAla 200
 DB 3303 TACTTCTCTGTGGTGGCTCAGTGAAGACTGGATCCGTGTTCTGTCGGCGCAGCGCT 3244
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220
 DB 3243 TTGTCCTACTTCTACATGGTGTCCGCTGGGGTGGCTACGTGTTTCATCATCAACCTGATA 3184
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 3183 CCCCTGCACGCTCTCTGACTGCTCATTTATGGCAGGTACTCGCGCGGCTCTGCTGACGAC 3124
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 3123 TACAGCAGCTTCTACATCTCTGGACTGCTGTTCTCCATGAGATCCCTTCGTGGGATTC 3064
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 3063 CAACCGATACGACAGCAGTGAACACATGCTGCGCTGGAGTGTGTTGTGCTCTCTATGGCC 3004
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 3003 GTGGCCACCTTTCGCCCATTTCCAGTCCGTGCTGTCGCGCAACGAGTTCGGAAGCTGTT 2944
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 2943 ATCGTGGCGGATGCTGCTGGTGGCGTGGGTGCTCTTTGTGGCCGTCTGCTGCTCACC 2884
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 2883 CTGGGCGTGTGGCCCGCTGGAGTGGAGCTTCTACTCGCTGGGATATCTGGCTAGCC 2824
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 2823 AAGATCCACATTCCTCATTCGCTGCGGAGCATCAGCCACCATCTGCTGCTGCTG 2764
 QY 361 PhePhePheAsnLeuHisIleLeuValCysAlaPheProValGlyValTyrTyrCysIle 380
 DB 2763 TTCTTCTTTTGTATTCACATCTCTGGTGTGGCTTCCAGTGGGAGTGTGGTACTGCATC 2704
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 2703 AAGCAGATCAACGACGAGCGGTTTCTGCTGCTGCTACGCCATCAGTGGCGTTACTTC 2644
 QY 401 AlaGlyValMetValValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 2643 GCTGT 2584
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 2583 GTGGCTTTTTCGGGACTGTTGGATGTGTTCTCTGCAAGAGGATTCGTCTAAGCGAATGGC 2524
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 DB 2523 ACAGCCATTAAGCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 2464
 QY 461 LeuTyrAspLys 464
 DB 2463 CTGTACGACAG 2452

RESULT 4
 ABL18224/c
 ID ABL18224 standard; DNA; 2953 BP.
 XX
 AC ABL18224;
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6145.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.

Thu Dec 16 16:25:14 2004

XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PA 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Claim 1; SEQ ID NO 6145; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at: ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 2953 BP; 924 A; 628 C; 648 G; 753 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 9,41e-204 Length: 2953
 Score: 208.00 Matches: 208
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 26.87% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-8 (1-774) X ABL18224 (1-2953)
 QY 1 MetAsnArgThrProLysMetLeuAsnSerIyeValAlaGlyTyrSerSerLeuLeuThr 20
 DB 626 ATGAATCGGACGCGAAGATGCTGAACAGCAAGGTGGCTGGCTACAGCCTAATCACC 567
 QY 21 PheAlaIleLeuLeuIleAlaTatPheLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 566 TTGCCATCTGTAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGCTATC 507
 QY 41 ArgPheGluSerIleLeuHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
 DB 506 CGTTTCGAGTCGATTATCCATGATTTGATCCGTGGTTCAACTACCGGCGCCACCGCTAC 447
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80
 DB 446 ATGGTGCGAATGTTGGTCAACTCTCACTGTTGACGAGCGCATGGTATCCG 387
 QY 81 LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 386 CTCGGCAGGATTGTGGCGGTACCGTCTATCCGGCGCTGATGATTACGTCGCGCGAATC 327
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 326 CATTGGCTGTCGACGACTCAACATACCGGTCTCATATTGTGACATCTCGTGTTCTCTG 267
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGluLeuTyr 140

DB 265 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACTCTCCACGACGAGCTGTGG 207
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 206 TCGCGGGCGCGCGCTCTTCGCCCGCAGCTTCATCGCCATCGTGGCTGATCATCAGT 147
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 146 AGTCTGCTGGCTGGATCGTACGATACGAGGGCATTCGCAATTCGCCCTCGAGTTACCC 87
 QY 181 TyrPheLeuTyrValArgSerValIleThrGlySerValPheThrSerAlaAlaAla 200
 DB 86 TACTTCTCTGGTGGCTCGCTCAGTGAAGACTGGATCGGTCTTGTGTCGCGCGAGCCCT 27
 QY 201 LeuSerTyrPheTyrMetValSer 208
 DB 26 TTGTCTACTTCTACATGGTGICC 3
 RESULT 5
 ABL18208/c
 ID ABL18208 standard; DNA; 2785 BP.
 XX AC ABL18208;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6097.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical; Gene; ds.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PA 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Claim 1; SEQ ID NO 6097; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at: ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 2785 BP; 795 A; 644 C; 604 G; 742 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8.3e-75 Length: 2785
 Score: 83.00 Matches: 83
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.72% Indels: 0

DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x ABL37128 (1-2785)

QY 603 SerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAsp 622
 DB 2784 TCTCTGACGTGGACTACGTTTGGTGATCTTTGGCGGTGATCGGCTATTTCTGGCGAT 2725

QY 623 AspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIle 642
 DB 2724 GATATCAACAAGTTCCTGTGGATGTCGGAATTCGTAGGAGAGCATCCCAAGACATT 2665

QY 643 LysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyValAsp 662
 DB 2664 AAGGAACGATTAATCTTTACCGACCGCGGTGAATTCAGGGTAGATGCCAAGGTGCTCCG 2605

QY 663 AlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeu 682
 DB 2604 GCCCTGCTCACTGCTTATGTACAAATTAGCTACTACGATTCGGGGAATTGAAGTTG 2545

QY 683 AspTyrArg 685
 DB 2544 GACTACAGG 2536

RESULT 6

ABL37128
 ID ABL37128 standard; cDNA; 433 BP.
 AC ABL37128;
 DT 08-APR-2002 (first entry)

Human colon tumour antigen polynucleotide SEQ ID NO:717.
 Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 Human colon tumour metastatic antigen; diagnosis; gene; ss.
 Homo sapiens.
 WO2001:96388-A2.
 20-DEC-2001.
 08-JUN-2001; 2001WO-US018557.
 09-JUN-2000; 2000US-0210899P.
 20-FEB-2001; 2001US-0270216P.
 (CORI-) CORIXA CORP.
 Jiang Y, Harlocker SL, Secrist H;
 WPI; 2002-114514/15.

Novel isolated colon tumor polynucleotide differentially expressed in colon tumor or colon metastatic tumor and polypeptides encoded by them, useful for inhibiting development of cancer in patient.
 Claim 1; SEQ ID NO 717; 105pp; English.

ABL36412 to ABL38645 represent human colon tumour antigen cDNA clones (1) which were isolated from human colon tumour and colon metastatic tumour cDNA libraries. (1) have cytostatic activity and can be used in vaccine production. (1) can be used for stimulating and/or expanding T cells specific for a tumour protein on contact with the T cells. They are also useful for inhibiting the development of cancer in a patient. (1) can be used as probes or primers for nucleic acid hybridisation, for preparing mutant species primers, or primers for use in genetic constructions. (1) can be used in the diagnosis of a colon tumour

Sequence 433 BP; 136 A; 68 C; 104 G; 123 T; 0 U; 2 Other;
 Alignment Scores:

Pred. No.: 2,83e-33 Length: 433
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-8 (1-774) x ABL37128 (1-433)

QY 551 AlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaSerArgThr 570
 DB 26 GCACGAGTAAATGCTTGGTGGGATTATGCTATCAGATAGCTGGAATGGCTAATAGAAT 85

QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 DB 86 ACCTTGGTGGATATTAACACCTGGGAATACACCCATAGCAGTGGTGGGAAAGCTATG 145

QY 591 SerSer 592
 DB 146 TCTTCT 151

RESULT 7

AAA43486
 ID AAA43486 standard; cDNA; 483 BP.
 AC AAA43486;
 DT 21-AUG-2000 (first entry)

Mouse secreted expressed sequence tag SEQ ID NO:61.
 Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
 expressed sequence tag; EST; probe; chemotactic; proliferative;
 immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 antiviral; antidiabetic; antistimulant; vulnerable; antiparkinsonian;
 antitumor; osteoprotective; neuroprotective; nontoxic; antiparasitic;
 cerebroprotective; anticoagulant; antidiabetic; allergic condition;
 autoimmune disorder; multiple sclerosis; allergic condition;
 insulin dependent diabetes; asthma; myeloid cell deficiency; vaccine;
 lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 central nervous system disorder; Alzheimer's disease; stroke;
 Parkinson's disease; Huntington's disease; coagulation disorder;
 haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
 infection; depression; psoriasis; ss.
 Mus musculus.
 WO200021991-A1.
 20-APR-2000.
 15-OCT-1999; 99WO-US024206.
 15-OCT-1998; 98US-0104436P.
 (GENY) GENETICS INST INC.
 Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 Merberg D, Treacy M, Bowman MR;
 WPI; 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted expressed sequence tags (SESTs), useful for treating various disorders such as autoimmune, infectious, and central nervous system disorders.
 Claim 1; Page 217; 803pp; English.

AAA43426 to AAA45925 represent specifically claimed secreted expressed sequence tags (SESTs), isolated from human, mouse, chicken and rat tissue sources. The SESTs can have a range of activities depending on the tissues they were isolated from. The activities include: chemotactic;

CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
CC antifungal; antiviral; antidiabetic; antiaesthetic; vulnery; antitumor;
CC osteopathic; neuroprotective; nootropic; antiparkinsonian; antipsoriatic;
CC cerebroprotective; anticonvulsant; and antidiabetic. The sESTs can be
CC used for gene therapy and in vaccines. The sESTs are useful as probes for
CC the identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma, myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention

XX SQ Sequence 483 BP; 123 A; 115 C; 128 G; 117 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,14e-33 Length: 483
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-8 (1-774) x AAA43486 (1-483)

QY 551 AlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
DB 196 GCCCGGTGATGCTGGTGGGACTACGCTATCAGATGCTGGATGCGCAACAGGACC 255
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
DB 256 ACTCTGTGGTAAACACACCTGGAAACACAGCCACATCGCACTGGTGGGAAAGCTATG 315
QY 591 SerSer 592
DB 316 TCTTCC 321

RESULT 8
AAC01967
ID AAC01967 standard; cDNA; 507 BP.
XX AAC01967;
AC AAC01967;
XX 06-OCT-2000 (first entry)
DT Human secreted protein 5' EST, SEQ ID NO: 1965.
DE Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX Gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX P-PSDB; AAG01961.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 1965; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dr primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX cDNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors

XX SQ Sequence 507 BP; 158 A; 85 C; 118 G; 146 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,28e-33 Length: 507
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-8 (1-774) x AAC01967 (1-507)

QY 551 AlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
DB 113 GCACGAGTAATGCTTGGTGGGATTATGCTATCAGATGCTGGAATGCTAATAAGACT 172
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
DB 173 ACCTTGTGGTAAACACCTGGAAACACAGCCACATCGCACTGGTGGGAAAGCTATG 232
QY 591 SerSer 592
DB 233 TCTTCT 238

RESULT 9
AAF93968
ID AAF93968 standard; DNA; 764 BP.
XX AAF93968;
AC AAF93968;
XX 23-MAY-2001 (first entry)
DT Primer specific for DNA encoding secretory/membrane protein SEQ ID 402.
DE Human; secretory protein; membrane protein; vaccine; gene therapy;
XX rheumatoid arthritis; diabetes; PCR primer; ss.
XX Synthetic.
XX EP1067182-A2.
XX 10-JAN-2001.
XX 07-JUL-2000; 2000EP-00114090.
XX 08-JUL-1999; 99JP-00194179.
XX 11-JAN-2000; 2000JP-00118775.
XX 02-MAY-2000; 2000JP-00183766.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX

XX AC AAH18021;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human cDNA sequence SEQ ID NO:17837.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX XX 07-FEB-2001.
 XX XX 28-JUL-2000; 2000EP-00116126.
 XX XX 29-JUL-1999; 99JP-00248036.
 XX XX 27-AUG-1999; 99JP-00300253.
 XX XX 11-JAN-2000; 2000JP-00118776.
 XX XX 02-MAY-2000; 2000JP-00183767.
 XX XX 09-JUN-2000; 2000JP-00241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 XX PS Claim 8; SEQ ID NO 17837; 2537pp + Sequence Listing; English.
 XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92445 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 XX SQ Sequence 2284 BP; 726 A; 388 C; 436 G; 734 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,39e-32 Length: 2284
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-8 (1-774) x AAH18021 (1-2284)
 QY 551 AlaArgValMetSerTTPAspTyrGlnIleAlaGlyMetAlaAsnArgThr 570

Db 177 GCACGAGTAATGCTCTGGTGGGATATGCTATCAATAGCTGGAATGGCTAATAGAACT 236
 QY 571 ThrLeuValAspAsnAsnThrTTPAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 Db 237 ACCTTGGTGGATAAATACACCTTGGATAACACCCATAGCACTGCTGGGAAAGCTATG 295
 QY 591 SerSer 592
 Db 297 TCCTCT 302
 RESULT 12
 ADD94783
 ID ADD94783 standard; cDNA; 2481 BP.
 XX AC ADD94783;
 XX DT 29-JAN-2004 (first entry)
 XX DE Human SIMP cDNA sequence.
 XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ss.
 XX OS Homo sapiens.
 XX PN W02003054008-A2.
 XX XX 03-JUL-2003.
 XX XX 18-DEC-2002; 2002WO-CA001967.
 XX XX 20-DEC-2001; 2001US-00028384.
 XX XX (COMP-) COMPATIGENE INC.
 XX XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 XX P-PSDB; ADD94784.
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.
 XX Claim 6; SEQ ID NO 1; 66pp; English.
 XX This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcoma, prostate cancer, testicular cancer, breast cancer, melanoma, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the human SIMP cDNA which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 1.49e-32 | Length: | 2481 |
| Score: | 42.00 | Matches: | 42 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 5.43% | Indels: | 0 |
| DB: | 10 | Gaps: | 0 |

US-10-028-384-8 (1-774) x ADD94783 (1-2481)

QY 551 AlaArgValMetSerTrrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

Db 1795 GCACGAGTAATGCTTGTGGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 1854

QY 571 ThrLeuValAspAsnAsnThrTrrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590

Db 1855 ACGTTGGTGGATAATACACCTGGATAACACCCACATAGCACTGGTGGGAAAGCTATG 1914

QY 591 SerSer 592

Db 1915 TCTTCT 1920

US-10-028-384-8 (1-774) x AAF93772 (1-2510)

QY 551 AlaArgValMetSerTrrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

Db 268 GCACGAGTAATGCTTGTGGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 327

QY 571 ThrLeuValAspAsnAsnThrTrrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590

Db 328 ACGTTGGTGGATAATACACCTGGATAACACCCACATAGCACTGGTGGGAAAGCTATG 387

QY 591 SerSer 592

Db 388 TCTTCT 393

RESULT 14

AAD08289

ID AAD08289 standard; cDNA; 2537 BP.

XX AC AAD08289;

XX DT 08-AUG-2001 (first entry)

XX DE Human secreted protein-encoding gene 7 cDNA clone HDTLR06, SEQ ID NO: 17.

XX KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma; foetal abnormality; developmental abnormality; haematopoietic disorder; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; Parkinson's disease; cognitive disorder; schizophrenia; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder; inflammation; neurological disorder; Alzheimer's disease; food additive; angiotensin-related disorder; kidney disorder; gastrointestinal disorder; allergy; pregnancy-related disorder; endocrine disorder; infection; wound healing; cell culture; chemotaxis; vulnerability; binding partner identification; gene therapy; ss.

XX OS Homo sapiens.

XX PH Key

XX FT CDS

XX FT 93..953

XX FT /*tag= a

XX FT /product= "Human secreted protein precursor"

XX FT sig_peptide

XX FT 93..155

XX FT /*tag= b

XX FT mat_peptide

XX FT 156..950

XX FT /*tag= c

XX FT /product= "Mature human secreted protein"

XX WO200136440-A1.

XX 25-MAY-2001.

XX

SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 1.49e-32 | Length: | 2481 |
| Score: | 42.00 | Matches: | 42 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 5.43% | Indels: | 0 |
| DB: | 10 | Gaps: | 0 |

US-10-028-384-8 (1-774) x ADD94783 (1-2481)

QY 551 AlaArgValMetSerTrrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

Db 1795 GCACGAGTAATGCTTGTGGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 1854

QY 571 ThrLeuValAspAsnAsnThrTrrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590

Db 1855 ACGTTGGTGGATAATACACCTGGATAACACCCACATAGCACTGGTGGGAAAGCTATG 1914

QY 591 SerSer 592

Db 1915 TCTTCT 1920

US-10-028-384-8 (1-774) x AAF93772 (1-2510)

QY 551 AlaArgValMetSerTrrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

Db 268 GCACGAGTAATGCTTGTGGGATATGGCTATCAGATAGCTGAATGGCTAATAGAACT 327

QY 571 ThrLeuValAspAsnAsnThrTrrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590

Db 328 ACGTTGGTGGATAATACACCTGGATAACACCCACATAGCACTGGTGGGAAAGCTATG 387

QY 591 SerSer 592

Db 388 TCTTCT 393

RESULT 13

AAF93772

ID AAF93772 standard; cDNA; 2510 BP.

XX AC AAF93772;

XX DT 23-MAY-2001 (first entry)

XX DE Human cDNA encoding a membrane or secretory protein clone PSECC0070.

XX KW Human; secretory protein; membrane protein; vaccine; gene therapy; rheumatoid arthritis; diabetes; ss.

XX OS Homo sapiens.

XX PF EPI067182-A2.

XX PD 10-JAN-2001.

XX PF 07-JUL-2000; 2000EP-00114090.

XX PR 08-JUL-1999; 99JP-00194179.

XX PR 11-JAN-2000; 2000JP-00118775.

XX PR 02-MAY-2000; 2000JP-00183766.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX DR WPI; 2001-093989/11.

XX DR P-PSDB; AAB88345.

XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in gene therapy or as candidate target molecules in drug development.

XX PS Claim 1; SEQ ID NO 57; 609pp + Sequence Listing; English.

XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916 which encode human secretory or membrane proteins represented by AAB88317 - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the invention. The invention also includes methods for the production of antibodies directed against the proteins, and cDNA sequences, which can be used in vaccines. The polynucleotide sequences can be used in gene therapy. The polynucleotide sequences and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect

CC and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). Examples of diseases which may be treated include rheumatoid arthritis and diabetes

us-10-028-384-8.Oligo.rng

Thu Dec 16 16:25:14 2004

PF 15-NOV-2000; 2000WO-US031282.
 XX
 XX
 PR 19-NOV-1999; 99US-0166414P.
 PR 21-JUL-2000; 2000US-0219665P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Komatsculis GA, Birse CE, Moore PA;
 XX WPI; 2001-343795/36.
 DR P-PSDB; AAE03824.
 XX
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX
 XX Claim 1; Page 443; 553pp; English.
 PS
 XX
 XX AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
 CC protein genes, and AAE03818-AAE03870 represent the proteins they encode.
 CC AAE03871-AAE03896 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 23 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g. rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g. Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g. psoriasis),
 CC sepsis, diabetes, atherosclerosis, gastrointestinal disorders, angio-
 CC genic disorders, kidney disorders, endocrine disorders, and infections. The
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
 CC present sequence represents a human secreted protein-encoding cDNA of the
 CC invention
 XX
 XX Sequence 2537 BP; 816 A; 434 C; 487 G; 800 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 1.52e-32 Length: 2537
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-8 (1-774) x AAD08289 (1-2537)
 QY 551 AlaArgValMetSerTrpTtpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 DB 267 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGATGGCTATAGACT 326
 QY 571 ThrLeuValAspAsnAsnThrTtpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 DB 327 ACGTTGGTGGATATAACACCTGGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATG 386
 QY 591 SerSer 592
 DB 387 TCCTCT 392
 RESULT 15
 AAV44866

AAV44866 standard; cDNA; 2546 BP.
 AAV44866;
 21-OCT-1998 (first entry)
 Clone CT585_1 coding sequence.
 Secreted protein; nutritional source; cell proliferation activity;
 cell differentiation activity; immune stimulant; tissue growth activator;
 haematopoiesis regulator; anti-inflammatory; tumour invasion suppressor;
 tumour inhibitor; clone CT585_1; ds.
 Homo sapiens.
 Location/Qualifiers
 112..972
 CDS /*tag= a
 WO9825962-A2.
 18-JUN-1998.
 12-DEC-1997; 97WO-US023224.
 13-DEC-1996; 96US-00766263.
 11-DEC-1997; 97US-00989232.
 (GEMY) GENETICS INST INC.
 Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
 Spaulding V, Agostino WJ;
 WPI; 1998-362424/31.
 P-PSDB; AAE9247.
 New isolated polynucleotides - obtained from human adult testis, human
 adult ovary, human adult brain and human adult heart cDNA libraries.
 Claim 35; Page 79-81; 108pp; English.
 This sequence represents a polynucleotide of the invention, and encodes a
 secreted protein. It was isolated from a human adult brain cDNA library,
 and is designated clone CT585_1. The DNA sequences and encoded
 polypeptides can be used as nutritional sources or supplements, or may
 exhibit e.g. cytokine and cell proliferation/differentiation activity,
 immune stimulating or suppressing activity, haematopoiesis regulating
 activity, receptor/ligand activity, anti-inflammatory activity,
 activin/inhibin activity, chemostatic/chemokinetic activity,
 cadherin/tumour invasion suppressor activity, tissue growth activity,
 tumour inhibition activity or other activities
 Sequence 2546 BP; 837 A; 416 C; 490 G; 803 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.53e-32 Length: 2546
 Score: 42.00 Matches: 42
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 5.43% Indels: 0
 DB: 2 Gaps: 0
 US-10-028-384-8 (1-774) x AAV44866 (1-2546)
 QY 551 AlaArgValMetSerTrpTtpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
 DB 286 GCACGAGTAATGCTTGGTGGGATTATGGCTATCAGATAGCTGGATGGCTATAGACT 345
 QY 571 ThrLeuValAspAsnAsnThrTtpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
 DB 346 ACGTTGGTGGATATAACACCTGGGATTAACAGCCACATAGCACTGGTGGGAAAGCTATG 405
 QY 591 SerSer 592

Db |||||
 406 TCTCT 411

Search completed: December 14, 2004, 23:35:02
Job time : 918.777 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31, Search time 974.477 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRTPKMLNSKVAGYSSLLT.....RRKGYIENRPVVGKRTLK 774

Scoring table:

Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat.14122004.131648.6892/app_query.fasta.1.3740
-DB=Published Applications NA -QFWT=fastap -SURFIX=Oligo.rnpb -MINMATCH=0.1
-LOPCU=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384 @CEN 1.1 2124 @runat.14122004.131648.6892
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESOURCY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 774 | 100.0 | 2417 | 15 | US-10-028-384-7 |
| 2 | 42 | 5.4 | 430 | 9 | US-09-878-178-717 |
| 3 | 42 | 5.4 | 430 | 13 | US-10-046-935-717 |
| 4 | 42 | 5.4 | 430 | 14 | US-10-146-502-717 |
| 5 | 42 | 5.4 | 1828 | 10 | US-09-945-527-62 |
| 6 | 42 | 5.4 | 2481 | 15 | US-10-028-384-7 |
| 7 | 42 | 5.4 | 2660 | 16 | US-10-264-059-630 |
| 8 | 42 | 5.4 | 2710 | 15 | US-10-028-384-3 |
| 9 | 34 | 4.4 | 387 | 16 | US-10-276-774-173 |
| 10 | 34 | 4.4 | 558 | 14 | US-10-052-283-433 |
| 11 | 34 | 4.4 | 1209 | 15 | US-10-106-698-330 |
| 12 | 34 | 4.4 | 1209 | 16 | US-10-264-237-412 |
| 13 | 34 | 4.4 | 1543 | 10 | US-09-974-879-133 |
| 14 | 34 | 4.4 | 1543 | 13 | US-09-305-736-133 |
| 15 | 34 | 4.4 | 1543 | 10 | US-09-818-683-133 |
| 16 | 34 | 4.4 | 1543 | 11 | US-09-818-683-133 |
| 17 | 34 | 4.4 | 1543 | 16 | US-10-621-401-133 |
| 18 | 32 | 4.1 | 1848 | 15 | US-10-128-714-1139 |
| 19 | 32 | 4.1 | 1969 | 15 | US-10-128-714-1139 |
| 20 | 32 | 4.1 | 2232 | 15 | US-10-128-714-1139 |
| 21 | 32 | 4.1 | 2603 | 15 | US-10-128-714-1139 |
| 22 | 32 | 4.1 | 3969 | 15 | US-10-128-714-1139 |
| 23 | 32 | 4.1 | 4603 | 15 | US-10-128-714-1139 |
| 24 | 30 | 3.9 | 743 | 10 | US-09-945-527-63 |
| 25 | 30 | 3.9 | 1114 | 16 | US-10-296-115-629 |
| 26 | 28 | 3.6 | 500 | 9 | US-09-998-598-1643 |
| 27 | 28 | 3.6 | 616 | 9 | US-09-879-536-332 |
| 28 | 27 | 3.5 | 474 | 14 | US-10-052-283-454 |
| 29 | 23 | 3.0 | 637 | 15 | US-10-002-631C-183 |
| 30 | 23 | 3.0 | 640 | 15 | US-10-002-631C-179 |
| 31 | 21 | 2.7 | 154 | 10 | US-09-991-936-450 |
| 32 | 21 | 2.7 | 2472 | 15 | US-10-171-581-112 |
| 33 | 21 | 2.7 | 2472 | 15 | US-10-028-384-11 |
| 34 | 21 | 2.7 | 2472 | 15 | US-10-172-118-742 |
| 35 | 21 | 2.7 | 2472 | 16 | US-10-342-887-742 |
| 36 | 21 | 2.7 | 3046 | 18 | US-10-417-375-95 |
| 37 | 21 | 2.7 | 3093 | 18 | US-10-417-375-92 |
| 38 | 21 | 2.7 | 3094 | 15 | US-10-028-384-9 |
| 39 | 21 | 2.7 | 5404 | 18 | US-10-417-375-99 |
| 40 | 21 | 2.7 | 5827 | 18 | US-10-417-375-97 |
| 41 | 20 | 2.6 | 393 | 16 | US-10-276-774-123 |
| 42 | 20 | 2.6 | 502 | 17 | US-10-767-701-6513 |
| 43 | 20 | 2.6 | 2157 | 9 | US-09-801-368-387 |
| 44 | 20 | 2.6 | 2157 | 18 | US-10-793-639-318 |
| 45 | 20 | 2.6 | 2681 | 16 | US-10-425-114-14408 |

ALIGNMENTS

RESULT 1
US-10-028-384-7
Sequence 7, Application US/10028384
Publication No. US20030148285A1
GENERAL INFORMATION:
APPLICANT: COMPATIGENE INC.
APPLICANT: PERREAU, Claude
APPLICANT: MCBRIDE, Kevin
TITLE OF INVENTION: Mammalian STMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028.384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 2417
TYPE: DNA
ORGANISM: Drosophila melanogaster
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: AF132552
DATABASE ENTRY DATE: 1999-04-27

: RELEVANT RESIDUES: (1) ... (2417)

US-10-028-384-7

Alignment Scores:

Alignment scores:
pred. No.:

pred. No.:
score.

Score: _____

Percent Correct: _____

Percent Similarity:

Best Local Similarity:

Query Match:

length: 2417

Length: 2417
Matches: 774

Matches: 774

Conservative: 0

Mismatches: 0

Indels: 0

UUS-10-028-384-8 (1-77A) y UUS-10-028-384-7 (1-2417)

| | | | |
|----|-----|--|-----|
| QY | 1 | MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerLeuLeuIleThr | 20 |
| Db | 71 | ATGAATCGGACGCCGAAGATCTGAACAGCAAGGTGGCTGGCTACAGCAGCCTAATCAC | 130 |
| QY | 21 | PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle | 40 |
| Db | 131 | TTGCGCATCTCTAATTCGCCTGGCTGGCGGGAATTTCTCTCGCCTCTTCGCGGTCA | 190 |
| QY | 41 | ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr | 60 |
| Db | 191 | CGTTTCGAGTCGATATCCATAGATTTGATCCGTGGTTCAACTACCGGCGCCACCGCTAC | 250 |
| QY | 61 | MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro | 80 |
| Db | 251 | ATGTGTCAGAATGGTTGGTACAACTTCCTCACTGGTTTCGACGAGCGCGCATGGTATCG | 310 |
| QY | 81 | LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle | 100 |
| Db | 311 | CTCGCAGGATTTGGGCGGTPACCGCTATATCCCGCGCTGATGATTACGTCCGCGCGGAATC | 370 |
| QY | 101 | HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu | 120 |
| Db | 371 | CATTGGCTGCTGCAGTACTCAACATACCGGTCATATTCGTGACATCTCGGTGTTCTGT | 430 |
| QY | 121 | AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIysGluLeuTrp | 140 |
| Db | 431 | GCGCCGATCTTCAGTGGCCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGGTGTGG | 490 |
| QY | 141 | SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer | 160 |
| Db | 491 | TCGCGGCGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCCTGGCTACATCAGT | 550 |
| QY | 161 | ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr | 180 |
| Db | 551 | AGGTCGGTGGCTGGATCGTACGATACAGAGGCAATGGCCATATTCGCCCTGCAGTTACCC | 610 |
| QY | 181 | TyrPheLeuTrpValArgSerValTyrGlySerValPheTrpSerAlaAlaAla | 200 |
| Db | 611 | TACTTCTGTGGGTGCGCTCAGTCAGACTGGATCCGTGTTCTGGTGGCGCGCAGCCGCT | 670 |
| QY | 201 | LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle | 220 |
| Db | 671 | TTGTCTCTACTTCTACATGGTGTCCGCTCGGGTGGCTACGTGTTCATCAACCTGATA | 730 |
| QY | 221 | ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer | 240 |
| Db | 731 | CCCTCTGCAGCTCTTCGTACTGCTCATTTATGGGAGGTTACTCGCGCGGTCTGCTGACCAGC | 790 |
| QY | 241 | TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe | 260 |
| Db | 791 | TACAGCACCTTCACATCTCGGAGTCTGTTCTTCATCGAGATCCCTTCGTGGGATTC | 850 |
| QY | 261 | GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuValAla | 280 |
| Db | 851 | CAACCGATACGACACGATGAACATGGCTGGCGCTGGAGTGTGTGTCTCTATGCCC | 910 |
| QY | 281 | ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe | 300 |
| Db | 911 | GTGGCCACCTTCGCGCAATTTGCAATCGTCTCTGTCCGCAAGAGATTCGCGAAGCTGTT | 970 |

Db 2051 GCTCGGCGCTGCTCACTGCTTATGACAAATTAAGCTACACTACAGATCGGGGATG 2110
QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
Db 2111 AAGTTGGACTACAGAGGTCCATCTGGATATGATCGCACAGTAACGCGTCATTGGGAAT 2170
QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrGluHisTyrPleuValArg 720
Db 2171 AAGGACTTCGATCTGACCTACTGGAGGAGGCGCTACCCAGACACTGCTGTTCGC 2230
QY 721 IleTyrArgValLysPheProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
Db 2231 ATCTATAGGCTGAAGAGCGCATGAGTTCAATAGACATCACTGGAAGACCAAGAGAGA 2290
QY 741 ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgTyrGlyTyrIle 760
Db 2291 AGATTTCCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTCCCAAGGCTACATA 2350
QY 761 ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774
Db 2351 CGAAACCGACCGTGTGTGTTAAGGGAACCAACCTTGAAA 2392

RESULT 2

US-09-878-178-717
; Sequence 717, Application US/09878178
; Patent No. US20020177552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-878-178-717

Alignment Scores:
Pred. No.: 7,59e-34 Length: 430
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-8 (1-774) x US-09-878-178-717 (1-430)

QY 551 AlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 26 GCACGAGTAATGCTCTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGACT 85
QY 571 ThrLeuValAspAsnAsnThrTyrPheAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 86 ACGTTGGTGATAATAACACCTCGGAATAACAGCCACATAGCACTGGTGGGAAAAGCTATG 145

QY 591 SerSer 592
Db 146 TCTTCT 151

RESULT 3

US-10-046-935-717
; Sequence 717, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.

; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-046-935-717

Alignment Scores:
Pred. No.: 7,59e-34 Length: 430
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 13 Gaps: 0

US-10-028-384-8 (1-774) x US-10-046-935-717 (1-430)

QY 551 AlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 26 GCACGAGTAATGCTCTGGTGGGATTATGGCTATCAGATAGCTGGAATGGCTAATAGACT 85
QY 571 ThrLeuValAspAsnAsnThrTyrPheAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 86 ACGTTGGTGATAATAACACCTCGGAATAACAGCCACATAGCACTGGTGGGAAAAGCTATG 145
QY 591 SerSer 592
Db 146 TCTTCT 151

RESULT 4

US-10-146-502-717
; Sequence 717, Application US/10146502
; Publication No. US20030069180A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aijun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C2
; CURRENT APPLICATION NUMBER: US/10/146,502
; NUMBER OF SEQ ID NOS: 2241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 717
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-146-502-717

Alignment Scores:
Pred. No.: 7,59e-34 Length: 430
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 14 Gaps: 0

US-10-028-384-8 (1-774) x US-10-146-502-717 (1-430)

QY 551 AlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570

```
Db      26 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGATGGCTAATAGAACT 85
Qy      571 ThrLeuValAspAsnAsnThrTrpAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||
Db      86 ACGTTGGTGGATAATAACACCTGGATAATACAGCCACATAGCACTGGTGGGAAAGCTATG 145
Qy      591 SerSer 592
        |||||
Db      146 TCTTCT 151

RESULT 5
US-09-945-527-62
; Sequence 62, Application US/09945527
; Publication No. US20030055588A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20030055588A1el Nucleic Acid Molecules Encoding
; FILE REFERENCE: Nucleic Acid and Protein Homologs
; CURRENT FILING DATE: 2001-08-29
; CURRENT APPLICATION NUMBER: US/09/945,527
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 1828
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1767..1771
; OTHER INFORMATION: n = A,T,C or G
US-09-945-527-62

Alignment Scores:
Pred. No.: 2,96e-33      Length: 1828
Score: 42.00           Matches: 42
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 5.43%      Indels: 0
DB: 10                 Gaps: 0

US-10-028-384-8 (1-774) x US-09-945-527-62 (1-1828)
Qy      551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
        |||||
Db      451 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGATGGCTAATAGAACT 510
Qy      571 ThrLeuValAspAsnAsnThrTrpAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||
Db      511 ACGTTGGTGGATAATAACACCTGGATAATACAGCCACATAGCACTGGTGGGAAAGCTATG 570
Qy      591 SerSer 592
        |||||
Db      571 TCTTCT 576

RESULT 6
US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

Alignment Scores:
Pred. No.: 3,94e-33      Length: 2481
Score: 42.00           Matches: 42
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 5.43%      Indels: 0
DB: 15                 Gaps: 0

US-10-028-384-8 (1-774) x US-10-028-384-1 (1-2481)
Qy      551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
        |||||
Db      1795 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGATGGCTAATAGAACT 1854
Qy      571 ThrLeuValAspAsnAsnThrTrpAsnSerHisIleAlaLeuValGlyLysAlaMet 590
        |||||
Db      1855 ACGTTGGTGGATAATAACACCTGGATAATACAGCCACATAGCACTGGTGGGAAAGCTATG 1914
Qy      591 SerSer 592
        |||||
Db      1915 TCTTCT 1920

RESULT 7
US-10-264-049-630
; Sequence 630, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 630
; LENGTH: 2660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2634)..(2634)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2637)..(2637)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-630

Alignment Scores:
Pred. No.: 4,21e-33      Length: 2660
Score: 42.00           Matches: 42
Percent Similarity: 100.00%  Conservatives: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 5.43%      Indels: 0
DB: 16                 Gaps: 0

US-10-028-384-8 (1-774) x US-10-264-049-630 (1-2660)
Qy      551 AlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
        |||||
Db      330 GCACGAGTAATGCTTGGTGGATTATGGCTATCAGATAGCTGGATGGCTAATAGAACT 389
Qy      571 ThrLeuValAspAsnAsnThrTrpAsnSerHisIleAlaLeuValGlyLysAlaMet 590
```

Db 390 ACCTGGTGGTAAATAACACCTTGGTAATACACCCACATGACCTGGTGGGAAGCTATG 449
QY 591 SerSer 592
Db 450 TCTTCT 455
RESULT 8
US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)...(2469)
US-10-028-384-3
Alignment Scores:
Pred. No.: 4,296-33 Length: 2710
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
DB: 15 Gaps: 0
US-10-028-384-8 (1-774) x US-10-028-384-3 (1-2710)
QY 551 AlaArgValMetSerTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 1862 GCCCGGGTCATGTCGTGGTGGACTACGGCTATCAGATGCTGGCATGCCAACAGNCC 1921
QY 571 ThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIleAlaMet 590
Db 1922 ACTCTGGTGGATAACAAACCTTGGAAACAGCCACATGCGCACTGGTGGGAAAAGCTATG 1981
QY 591 SerSer 592
Db 1982 TCTTCC 1987
RESULT 9
US-10-276-774-173
; Sequence 173, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 173
; LENGTH: 387
; TYPE: DNA

; ORGANISM: Homo sapiens
US-10-276-774-173
Alignment Scores:
Pred. No.: 1,628-25 Length: 387
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 16 Gaps: 0
US-10-028-384-8 (1-774) x US-10-276-774-173 (1-387)
QY 325 AlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
Db 235 GCACCATGGAGTGGCAGGCTTTTATTCATGTGGGATGCTGGGTATGCAAAATACACATT 294
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrTrp 358
Db 295 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 336
RESULT 10
US-10-052-283-433
; Sequence 433, Application US/10052283
; Publication No. US20030064379A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Goddard, Audrey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND METHOD OF USE THEREOF
; FILE REFERENCE: P2751R1C1
; CURRENT APPLICATION NUMBER: US/10/052,283
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: PCT/US00/20006
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/145,701
; PRIOR FILING DATE: 1999-07-26
; NUMBER OF SEQ ID NOS: 564
; SEQ ID NO 433
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 23, 111, 130, 146
; OTHER INFORMATION: unknown base
US-10-052-283-433
Alignment Scores:
Pred. No.: 2,296-25 Length: 558
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 14 Gaps: 0
US-10-028-384-8 (1-774) x US-10-052-283-433 (1-558)
QY 325 AlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
Db 224 GCACCATGGAGTGGCAGGCTTTTATTCATGTGGGATGCTGGGTATGCAAAATACACATT 283
QY 345 ProIleIleAlaSerValSerGluHisGlnProThrTrp 358
Db 284 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 325
RESULT 11
US-10-106-698-330
; Sequence 330, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides

```

; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 330
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-330

Alignment Scores:
Pred. No.: 4,73e-25 Length: 1209
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4,39# Indels: 0
DB: 15 Gaps: 0

US-10-028-384-8 (1-774) X US-10-106-698-330 (1-1209)
QY 325 AlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
DB 866 GCACCATGGAGTGGCAGGTTTTTATTCATTGTGGGATCTGGGTATGCAAAAATACACATT 925

RESULT 12
US-10-264-237-412
; Sequence 412, Application US/10254237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 412
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-237-412

Alignment Scores:
Pred. No.: 4,73e-25 Length: 1209
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4,39# Indels: 0
DB: 16 Gaps: 0

US-10-028-384-8 (1-774) X US-10-264-237-412 (1-1209)
QY 325 AlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIle 344
DB 866 GCACCATGGAGTGGCAGGTTTTTATTCATTGTGGGATCTGGGTATGCAAAAATACACATT 925

RESULT 13
US-09-974-879-133
; Sequence 133, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066,094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,089
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,095
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,090
; PRIOR FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-974-879-133

Alignment Scores:
Pred. No.: 5,95e-25 Length: 1543
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4,39# Indels: 0
DB: 16 Gaps: 0

```

Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-974-879-133 (1-1543)

QY 325 AlaProTtpSerGlyArgPheTyrSerLeuTtpAspThrGlyTyrAlaLysIleHisIle 344
|||
Db 866 GCACCATGGAGTGGCAGGTTTATTTCATTGTGGTACTGGTATGCAAAATACACATT 925
|||

QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTtp 358
|||
Db 926 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 967
|||

RESULT 14

US-09-305-736-133
; Sequence 133, Application US/09305736
; Publication No. US20030088078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; EARLIER FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-305-736-133

Alignment Scores:

Pred. No.: 5,95e-25 Length: 1543
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-305-736-133 (1-1543)

QY 325 AlaProTtpSerGlyArgPheTyrSerLeuTtpAspThrGlyTyrAlaLysIleHisIle 344
|||
Db 866 GCACCATGGAGTGGCAGGTTTATTTCATTGTGGTACTGGTATGCAAAATACACATT 925
|||

QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTtp 358
|||
Db 926 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 967
|||

RESULT 15

US-09-818-683-133
; Sequence 133, Application US/09818683
; Publication No. US20030211472A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: P2020P1
; CURRENT APPLICATION NUMBER: US/09/818,683
; CURRENT FILING DATE: 2001-03-28
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1055)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-818-683-133

Alignment Scores:

Pred. No.: 5,95e-25 Length: 1543
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.39% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x US-09-818-683-133 (1-1543)

QY 325 AlaProTtpSerGlyArgPheTyrSerLeuTtpAspThrGlyTyrAlaLysIleHisIle 344
|||
Db 866 GCACCATGGAGTGGCAGGTTTATTTCATTGTGGTACTGGTATGCAAAATACACATT 925
|||

QY 345 ProIleIleAlaSerValSerGluHisGlnProThrThrTtp 358
|||
Db 926 CCAATTATTGCATCAGTGTCTGAGCATCAACCTACGACTTGG 967
|||

Search completed: December 15, 2004, 15:27:41
Job time : 993.477 secs

Blank
Page

QY 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
 DB 202 CGTTTCGAGTCGATTCATCATGAGTTTGATCCGGTGGTTCAACTACCGGGCCCGCTAC 261
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrPro 80
 DB 262 ATGGTGCAGAAATGGTGGTGAACCTTCCTCAACTGGTTCGACGAGCGCATGGTATCCG 321
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 322 CTCGGCAGGATTTGGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATC 381
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 382 CATTTGGCTGCTGCACGCTACTCAACATACCGGTTCATATTCGTGACATCTGCGTGTTCCTG 441
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGlyLeuTyr 140
 DB 442 GCGCGGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 501
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 502 TCCGGCGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCTGCTGGCTGCTGCTGCTGCTG 561
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 562 AGGTCGGTGGCTGATCGTACGATAACGAGGCGATTCGCCATATTCGCCCTGCGATTCACC 621
 QY 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheThrSerProArgLeuLeuThrSer 240
 DB 622 TACTTCCCTGGGGCGCGCTCAGTGAAGACTGGATCCGTGTCTGTGTCGGCGCGCGCGCT 681
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
 DB 682 TTGCTCTACTTCTACATGGTGTCCGCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 741
 QY 221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 742 CCCCTGCACGCTTTCGTAAGTCTCATTTATGGGAGGTAATCGCGCGGTCTGCTGACCGAC 801
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 802 TACAGCACCTTCTACATCTCCGGGACTGCTGTCTCCATGCAGATCCCTTCGTGGGATTC 861
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGly 273
 DB 862 CAACCGATACGACCACTGAACACATGGCTGCGCTGGGA 900

RESULT 2

US-09-270-767-158
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 158
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Alignment Scores:

| | | | |
|------------------------|-----------|---------------|------|
| Pred. No.: | 1,26e-241 | Length: | 1660 |
| Score: | 246.00 | Matches: | 246 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 31.78% | Indels: | 0 |
| DB: | 4 | Gaps: | 0 |

US-10-028-384-8 (1-774) x US-09-270-767-158 (1-1660)

QY 529 SerArgAsnIleLeuAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla 548
 DB 788 TCCCGCAACATTTTAGCATTTTCAGAGAGGCTTACTACTGGCTTTCGCAAAACACGCC 847
 QY 549 AspAspAlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
 DB 848 GATGATGCTCGCGTTATGCTTGGTGGGATACGGATACCGATAGCGGGAATGGCAAC 907
 QY 569 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIys 588
 DB 908 AGAACACCGCTAGTGGATAATAATACGTGGAACAATAGTACATAGCGCTGGTGGCAAG 967
 QY 589 AlaMetSerSerThrGluGlyLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
 DB 968 GCAATGCTTCAACCGAGGAGAAGTCTTACGAAATTAATGATCATCTCTTGACGTGGACTAC 1027
 QY 609 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
 DB 1028 GTTTTGGTATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCCCTG 1087
 QY 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
 DB 1088 TGGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAGGACATTAAGGAAAGCGATTACTTT 1147
 QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
 DB 1148 ACCGACCGCGGTGAATTCAGGGTAGATCCGAAGGTGCTCCGGCCCTGCTCAACTGCGCTT 1207
 QY 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 689
 DB 1208 ATGTACAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTGGACTACAGAGGTCCATCT 1267
 QY 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
 DB 1268 GGATATGATCGCACACGTAACCGCTCATTTGGGAATTAAGGACTTCGATCTGACCTACCTG 1327
 QY 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
 DB 1328 GAGGAGCGCTACACACAGAACACTGGCTTGTTCGCATCTATAGGTGAAGAAGCCGCAT 1387
 QY 729 GluPheAsnArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIle 748
 DB 1388 GAGTTCAATAGACCATCACTGAAGACCAAGAGAGAACGATTCCTCCAGCAAACTTCATT 1447
 QY 749 SerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProValValValLys 768
 DB 1448 TCGAGAAAGAACTTAAGCGTCGCAAGGCTACATACGAAACCGCGGTGTTGTTAAG 1507
 QY 769 GlyLysArgThrLeuLys 774
 DB 1508 GGAAACGAACTTGAAA 1525

RESULT 3

US-09-270-767-15440
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

Alignment Scores:
 Pred. No.: 1.26e-241 Length: 1660
 Score: 246.00 Matches: 246
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.78% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)

| | | | |
|----|------|---|------|
| QY | 529 | SerArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAla | 548 |
| DB | 788 | TCCCGCAACATTTAGACGATTCAGAGAGGCTTACTAGCTTTCGCAAGACACTGCC | 847 |
| QY | 549 | AspAspAlaArgValMetSerTrpAspTyrGlyTrpGlnIleAlaGlyMetAlaAsn | 568 |
| DB | 848 | GATGATGCTCGGTTATGCTTTGGTGGATTCAGGATACCATAGCGGAAATGGCAAC | 907 |
| QY | 569 | ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisLeuAlaLeuValGlyLys | 598 |
| DB | 908 | AGAACGACGCTAGTGGATAAATACGTTGGAACAATAGTCACATAGCGCTGGTTGCCAAG | 967 |
| QY | 599 | AlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr | 608 |
| DB | 968 | GCAATGCTTCAACCGAGAGAGAGTCCCTACGAAATATGACATCTCTTGACGTGGACTAC | 1027 |
| QY | 609 | ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu | 628 |
| DB | 1028 | GTITTTGGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCCTG | 1087 |
| QY | 629 | TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe | 648 |
| DB | 1088 | TGGATGGTCCGAATTCCTGAGGAGAGCATCCCAAGGACATTAAGGAACGCAATTACTTT | 1147 |
| QY | 649 | ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu | 668 |
| DB | 1148 | ACCGACCGCGTGAATTCAGGCTAGATCCGAGGTGCTCGGCGCTCTCAACTGCTT | 1207 |
| QY | 669 | MetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer | 688 |
| DB | 1208 | ATGTACAATTAAGCTACTACAGATTCGGGAATTAAGTTGGACTACAGAGTCCATCT | 1267 |
| QY | 689 | GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu | 708 |
| DB | 1268 | GGATATGATCGCACACGTAAACCGCTCATTTGGGAATAAGGACTTCGATCTGACCTACCTG | 1327 |
| QY | 709 | GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis | 728 |
| DB | 1328 | GAGGAGGCTACACCAAGAACACTGGCTTGTTCGATCTATAGGGTGAAGAGCGCAT | 1387 |
| QY | 729 | GluPheAsnArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIle | 748 |
| DB | 1388 | GAGTTCAATAGACCATCACTGAAGACCAAGGAGAGACGATTCCTCCAGCAAACTTCATT | 1447 |
| QY | 749 | SerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnAspProValValLys | 768 |
| DB | 1448 | TCGAGAAAGAACTCTAAGCGTCGCAAGGGCTACATACAGAAACCGCGTTGTGTGAAG | 1507 |
| QY | 769 | GlyLysArgThrLeuLys 774 | |
| DB | 1508 | CGAAAACGAACCTTGAAA 1525 | |

RESULT 4
 US-09-270-767-1318/c
 ; Sequence 1318, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 1318
 LENGTH: 560
 TYPE: DNA
 ORGANISM: Drosophila melanogaster
 US-09-270-767-1318

Alignment Scores:
 Pred. No.: 3.42e-156 Length: 560
 Score: 162.00 Matches: 162
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.93% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-1318 (1-560)

| | | | |
|----|-----|--|-----|
| QY | 369 | ValCysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgVal | 388 |
| DB | 492 | GTGTGCGCTTCCAGTGGAGGTGTGTACTGATCAAGCAGATCAACGAGCGCGTT | 433 |
| QY | 389 | PheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMet | 408 |
| DB | 432 | TTCTGTGTCTGTACGCCATCAGTCGGTTTACTTCGCTGGTGTGATGGTGGTTTATG | 373 |
| QY | 409 | LeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAsp | 428 |
| DB | 372 | TTGACCTCTACGCGGTGTGTGATGTCGCGGAGTGGCTTTTCGGGACTGTTGGAT | 313 |
| QY | 429 | ValPheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGlu | 448 |
| DB | 312 | GTGTTCCTGCAAGAGGATTCGTCTAGCGAATGGCACAGCCATAAGCGCAGCCAGAA | 253 |
| QY | 449 | ValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu | 468 |
| DB | 252 | GTGATGAGCTGAGGATTCATTTGAGAAGAAGCGCTGTACCAAGGCTGCAAGCTG | 193 |
| QY | 469 | LysHisArgThrLysHisAspAlaGlnAspThrGlyValSerSerAsnLeuLysSer | 488 |
| DB | 192 | AAGCATCGTACTAAGCATGATGCCACAGGATCTGGCGTCAGCTCCCAACCTGAAGAGT | 133 |
| QY | 489 | IleValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpVal | 508 |
| DB | 132 | ATTGTTATTTGGCGCTTCTAATGCTGTGATGATGTCCTGCTCCACTGCAGTGGTG | 73 |
| QY | 509 | ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly | 528 |
| DB | 72 | ACCAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTCCACCAACAGTCAGATGA | 13 |
| QY | 529 | SerArg 530 | |
| DB | 12 | TCCCGC 7 | |

RESULT 5
 US-09-270-767-16600/c
 ; Sequence 16600, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16600
 ; LENGTH: 560
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-16600

Alignment Scores:
 Pred. No.: 3.42e-156 Length: 560

Score: 162.00 Matches: 162
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 20.93% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-16600 (1-560)

QY 369 ValCysAlaPheProValGlyValTrpTyrCysIleIysGlnIleAsnAspGluArgVal 388
 DB 492 GTGTGGCGCTCCAGTGGAGTGTGTACTGATCAGCAGATCAACGACGAGCGCTT 433
 QY 389 PheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMet 408
 DB 432 TTCGTGTGTGTACGGCCATCAGTGGCGTTTACTTCGCTGTGTGTGTGTGTGTGTGT 373
 QY 409 LeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAsp 428
 DB 372 TTGACCTTCACCGCGT 313
 QY 429 ValPheLeuGlnAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGlu 448
 DB 312 GTGTTCCTGCAAGAGGATTCGTGTAGCGAATGGGCACACGATTAAGCGCAGCCGAA 253
 QY 449 ValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu 468
 DB 252 GTGATGAGCTGAGGATTCATTTGAGAGAGAGACCTGTACGACAAAGCTGGCAAGCTG 193
 QY 469 LysHisArgThrLysHisAspAlaGlnInAspThrGlyValSerSerAsnLeuLysSer 488
 DB 192 AAGCATCGTACTAAGCATGATGCGCCAGCAGGATCTGGCGTCCAGCTCCCAACCTGAAGAT 133
 QY 489 IleValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpVal 508
 DB 132 ATTGTATTGGCGGTTCTAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 73
 QY 509 ThrSerAsnAlaTyrSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
 DB 72 ACCAGCAATGCTACTCCAGTCCCTCCATGTCTTGCTTTCACAAACAGTCAAGATGGA 13
 QY 529 SerArg 530
 DB 12 TCCCGC.7

RESULT 6

US-09-270-767-158/c
 ; Sequence 158, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 158
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-158

Alignment Scores:
 Pred. No.: 1,15e-152 Length: 1660
 Score: 159.00 Matches: 210
 Percent Similarity: 99.06% Conservative: 0
 Best Local Similarity: 99.06% Mismatches: 1
 Query Match: 20.54% Indels: 2
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-158 (1-1660)

QY 564 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 583

DB 793 GCGGGATGGCAACAGACGACGCTAGTGGATATAATACGTGGACATATGTCACATA 734
 QY 584 AlaLeuValGlyIleAlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSer 603
 DB 733 GCGCTGCTTGGCAAGCAATGTCTTCAACCGAGGAGAGAGTCTCTACGAAATATGACATCT 674
 QY 604 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 623
 DB 673 CTTCAGCTGGATACGTTTGGTGTATCTTTGGCGGTGTATCGGCTATTCGGGCGATGAT 614
 QY 624 IleAsnLysPheLeuTrpMetValArgIleAlaGluGlyHisProLysAspIleLys 643
 DB 613 ATCAACAAGTCTCTGTGGATGTCGGAATTCGCGAGGAGAGCATCCCAAGGACATTAAG 554
 QY 644 GluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAla 663
 DB 553 GAAAGCGATATCTTACCGACCGCGTGTATTCAGGGTGTATGCGGAGGTGCTCCGGCC 494
 QY 664 LeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAsp 683
 DB 493 CTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGAATTCGAGTTGGAC 434
 QY 684 TyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 DB 433 TACAGAGGCCCATCTGGATGTATGTCACAGTAAACGCGTCATTCGGATTAAGGACTTT 374
 QY 704 AspLeuThrTyrLeuGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
 DB 373 GATCTGACCTACCTGGAGGAGGCTTACACCAAGCAACTGGCTTGTTCGCATCTATAG- 315
 QY 724 -VallyLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePr 743
 DB 314 TGTGAAGAAGCCGATGATTCATAGACCATCACTGAGACCCAGGAGAGACCATTC 255
 QY 743 oProAlaAsnPhelIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnAr 763
 DB 254 TCCAGCAAACTTCATCTCGAAGAACTCGAAGCGTCGCAAGGGCTACATACGAAACCG 195
 QY 763 gProValValLysGlyLysArgThrLeuLys 774
 DB 194 ACCGTTGTGTTAAGGGAAACGAAACCTTGAAA 161

RESULT 7

US-09-270-767-15440/c
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15440
 ; LENGTH: 1660
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440

Alignment Scores:
 Pred. No.: 1,15e-152 Length: 1660
 Score: 159.00 Matches: 210
 Percent Similarity: 99.06% Conservative: 0
 Best Local Similarity: 99.06% Mismatches: 1
 Query Match: 20.54% Indels: 2
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)

QY 564 AlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIle 583

Db 793 GCGGGAATGGCAACAGACGCTAGTGTGATATATATACGTGGAAACATAGTACATA 734
QY 584 AlalaValGlyLysAlaValSerThrGluGluLysSerThrGluLeuMetThrSer 603
Db 733 GCGCTGGTGGCAAGCAATGTCTTCAACCGAGGAGAGTCTTACGAAATTTATGACATCT 674
QY 604 LeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAsp 623
Db 673 CTGACGTGGACTAGCTTTGGTGTATCTTTGGCGGTGTGATCGGCTATTTCCGGCGATGAT 614
QY 624 IleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIleLys 643
Db 613 ATCAACAAGTTCTCTGTGATGGTCCGAATTCGAGGAGAGAGATCCCAAGGACATTAG 554
QY 644 GluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAla 663
Db 553 GAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGATGCGGAAGGTGCTCCGGCC 494
QY 664 LeuLeuAsnLysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAsp 683
Db 493 CTGCTCAACTGCCTTATGTACAATTAAGCTACTACAGATTCGGGGATTTGAAGTTGGAC 434
QY 684 TyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
Db 433 TACAGAGCCCATCTGGATATGATCGCACACGTAAACCGCGTCAATTGGGAATAAGGACTTT 374
QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
Db 373 GATCTGACCTACTGGAGGAGGCTTACACACAGAACACTGGCTTGTTCGCATCTATAG- 315
QY 724 -ValLysIleProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePr 743
Db 314 TGTGAAGAAGCGCATGAGTTCAATAGACCATCACTAGACCAAGAGAGAGACGATTCC 255
QY 743 oProAlaAsnPhelIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnAr 763
Db 254 TCCAGCAAACTTCTCTCGAAGAAAGAACTCGAAGCGTCCGAGGGCTTACATACGAACCG 195
QY 763 gProValValLysGlyLysArgThrLeuLys 774
Db 194 ACCGGTTGTTGTTAAGGGAACGAACTTGAAC 161
RESULT 8
US-09-270-767-28082
; Sequence 28082, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28082
; LENGTH: 133
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-28082
Alignment Scores:
Pred. No.: 3 85e-36 Length: 133
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.68% Indels: 0
Gaps: 4
US-10-028-384-8 (1-774) x US-09-270-767-28082 (1-133)
QY 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 2 ATGGGAGGAGTACTCGCGCGCTGTGCTGACCACTAGCAGCACCTTCTACATCTCTGGGACTG 61

QY 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 62 CTGTTCTCCATGCAGATCCCTTCGTGGGATTCACACCGATAGCCACCATGACACATG 121
QY 270 AlaAlaLeuGly 273
Db 122 GCTGCGCTGGGA 133
RESULT 9
US-09-513-999C-1965
; Sequence 1965, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.in
; SEQ ID NO 1965
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; FEATURE:
; LOCATION: 122..505
US-09-513-999C-1965
Alignment Scores:
Pred. No.: 1.57e-33 Length: 507
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.43% Indels: 0
Gaps: 4
US-10-028-384-8 (1-774) x US-09-513-999C-1965 (1-507)
QY 551 AlaArgValMetSerTrpTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThr 570
Db 113 GCACAGTAATGTCTTGTGGGATTTATGGCTATCAGATAGCTGGAATGGCTAATAGAACT 172
QY 571 ThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMet 590
Db 173 ACGTTGGTGGATATAACACCTGGAATAACACGACCATAGCCTGGTGGGAAAGCTATG 232
QY 591 SerSer 592
Db 233 TCTTCT 238
RESULT 10
US-09-328-111-332
; Sequence 332, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.

QY 151 PheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 170
DB 445 TTTATAGCCATTGTCCTCCGGTTATATCTAGATCAGTGGCGGGTCTCTACGATAATGAG 504

RESULT 14

US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:
Pred. No.: 5,128-10 Length: 503
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-27262 (1-503)

QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePhe 175
DB 2 CGGTATATCTCTCGATCCGTGGCGGATCGTACGACAAATGAAGGCAATCGGCATTTTC 58

RESULT 15

US-09-248-796A-3089
; Sequence 3089, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3089
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (25)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3089

Alignment Scores:
Pred. No.: 8,688-10 Length: 867
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2.45% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-248-796A-3089 (1-867)

QY 618 GlyTy-SerGlyAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGly 636

DB 433 GGTATTCTGGTGGATGATTAACAAATTTCTATGGATGGTAAGAATTGCTGAAGGT 489

Search completed: December 14, 2004, 23:43:55
Job time : 161.806 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:58 : Search time 5991.52 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 774

Sequence: 1 MNRTPKMLNKKVAGYSSLT.....RRKGVIIRNPVWVGKRTLK 774

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Word size: 1

Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgnt2_1/USFTO.spool/US10028384/runat_14122004_131646_6807/app_query.fasta_1.3740
-DB=EST -Q=MT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPEXT=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US10028384 @CGN 1 13960 @runat_14122004_131646_6807 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -Delop=6 -Delext=7

Database :

EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_est4:
5: gb_est5:
6: gb_est6:
7: gb_gss1:
8: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 245 | 31.7 | 810 | 2 | BF502026 |
| 2 | 221 | 28.6 | 701 | 2 | BF503420 |
| 3 | 214 | 27.6 | 655 | 2 | BF496296 |
| 4 | 214 | 27.6 | 715 | 7 | CK657451 |
| 5 | 208 | 26.9 | 679 | 7 | CK657649 |
| 6 | 208 | 26.9 | 681 | 7 | CK659033 |
| 7 | 197 | 25.5 | 652 | 4 | BI171940 |
| 8 | 196 | 25.3 | 666 | 4 | BI357074 |
| 9 | 196 | 25.3 | 668 | 4 | BI227902 |

| | | | | | |
|----|-----|------|-----|---|----------|
| 10 | 195 | 25.2 | 641 | 1 | AI257750 |
| 11 | 195 | 25.2 | 663 | 4 | BG641172 |
| 12 | 194 | 25.1 | 616 | 2 | BF504378 |
| 13 | 191 | 24.7 | 630 | 1 | AI295381 |
| 14 | 191 | 24.7 | 644 | 4 | BG636414 |
| 15 | 191 | 24.7 | 654 | 4 | BI374334 |
| 16 | 189 | 24.4 | 836 | 9 | CNS0172X |
| 17 | 188 | 24.3 | 642 | 4 | BI173041 |
| 18 | 187 | 24.2 | 635 | 1 | AI135629 |
| 19 | 184 | 23.8 | 600 | 4 | BI163592 |
| 20 | 183 | 23.6 | 736 | 1 | AA949890 |
| 21 | 182 | 23.5 | 694 | 4 | BI354111 |
| 22 | 180 | 23.3 | 615 | 4 | BI364613 |
| 23 | 179 | 23.1 | 555 | 4 | BI484774 |
| 24 | 178 | 23.0 | 598 | 4 | BI170396 |
| 25 | 176 | 22.7 | 585 | 1 | AI133998 |
| 26 | 174 | 22.5 | 564 | 4 | BI635692 |
| 27 | 170 | 22.0 | 658 | 4 | BI374189 |
| 28 | 168 | 21.7 | 585 | 4 | BI242003 |
| 29 | 167 | 21.6 | 539 | 4 | BI373819 |
| 30 | 167 | 21.6 | 564 | 1 | AI542327 |
| 31 | 166 | 21.4 | 649 | 4 | BI162865 |
| 32 | 162 | 20.9 | 690 | 4 | BG640860 |
| 33 | 160 | 20.7 | 554 | 4 | BI167839 |
| 34 | 159 | 20.5 | 555 | 4 | BI370013 |
| 35 | 155 | 20.0 | 646 | 4 | BI170481 |
| 36 | 154 | 19.9 | 525 | 2 | BF502679 |
| 37 | 148 | 19.1 | 539 | 1 | AI107144 |
| 38 | 147 | 19.0 | 495 | 1 | AI533215 |
| 39 | 147 | 19.0 | 527 | 4 | BI586931 |
| 40 | 146 | 18.9 | 441 | 7 | CO295609 |
| 41 | 141 | 18.2 | 424 | 7 | CO337666 |
| 42 | 140 | 18.1 | 502 | 4 | BI237763 |
| 43 | 138 | 17.8 | 414 | 7 | CO295621 |
| 44 | 138 | 17.8 | 414 | 7 | CO304185 |
| 45 | 138 | 17.8 | 414 | 7 | CO312883 |

ALIGNMENTS

BF502026 810 bp mRNA linear EST 19-APR-2001
AT17637.5prime AT Drosophila melanogaster adult testes POFB7
Drosophila melanogaster cDNA clone AT17637 5 similar to OatStt3:
FBan0007748 'enzyme' located on: 3R 96B16-96B17: 04/09/2001, mRNA
sequence.

ACCESSION BF502026 GI:13695997

VERSION BF502026.2

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 810)

AUTHORS

Stapleton, M., Brokstein, P., Hong, J., Agbayani, A., Baxter, E.,
Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N.,
Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nuncio, J.,
Paclet, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
Lewis, S. E., Celniker, S. and Rubin, G. M.
BDGS/HMI AT Drosophila EST Project

TITLE

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi11585327.

COMMENT

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd,

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AE003569: arm.X [20544566-20835979]

estimated-cyclo:19E4-19F1: 04/09/2001 hit P element 1(3)j299:

1(3)j2D9 AQ026308 inserted at base 292 5' end of P element Inverse
 PCR: 02/10/2001
 Plate: AT.176 row: D column: 1
 High quality sequence stop: 713.
 Location/Qualifiers

FEATURES

source

1..810
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="AT11637"
 /sex="male"
 /dev_stage="0-3 day old Ore-R males"
 /lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
 AT.121-AT.319: DHS-alpha Tona"
 /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
 /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
 Site_2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into pOTB7. Plasmid cDNA library."

ORIGIN

Alignment Scores:
 Pred. No.: 2,968-252 Length: 810
 Score: 245.00 Matches: 245
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.65% Indels: 0
 DB: 2 Gaps: 0

US-10-028-384-8 (1-774) x BF502026 (1-810)

Qy 1 MetAsnArgThrProLyMetLeuAsnSerLySValAlaGlyTyrSerSerLeuLeuThr 20
 Db 49 ATGAATCGGACCGCGGAGATCTCTGAACAGCAAGTGGCTGGCTACAGCGCTATACACC 108
 Qy 21 PheAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
 Db 109 TTCGGCATCTCTGATCGCTGGCTGGCGGATTTTCCTCTGCTCTCTGCGGTCATC 168
 Qy 41 ArgPheGluSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 60
 Db 169 CGTTTCGATGCGATTCATCATGATGTTGATCCGCGTGTCAACACCGCGGCCACCGCTAC 228
 Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80
 Db 229 ATGGTGCAGATGGTGTGTACACTCTCTCACTGTTCCAGCGCGGATGATATCCG 288
 Qy 81 LeuGlyArgTleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 289 CTCGCGAGATGTGGCGGTACCGTCTATCCCGCGCTGATGATTACGTCCGCGGAATC 348
 Qy 101 HisTyrLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 349 CATTCGCTGCTGCAGTACTCAACATACCGTCTCATATTCGTGACATCTCGGTGTTCTTG 408
 Qy 121 AlaProllePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGluLeuThr 140
 Db 409 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 468
 Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 469 TCCGCGGCGCGCGCTCTTCGCGCGCATCTTCATCGCCATCGTGGCTGCTACATCACT 528
 Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 529 AGGTGGGTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 588
 Qy 181 TyrPheLeuTyrValArgSerValIleThrGlySerValPheThrSerAlaAlaAlaAla 200
 Db 589 TACTTCCGTGGTGGCTCAGTGAAGACTGATCCGCTGTTCTGCTGGCGCGGCGCT 648

Qy 201 LeuSerTyrPheTyrMetValSerAlaTyrGlyGlyTyrValPheIleLeuLeuLeu 220
 Db 649 TTATCTTACTTCTACATGTTGTCGCCCTGGGTGGCTACGTGTTCATCATCACTGATA 708
 Qy 221 ProLeuHisValPheValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 240
 Db 709 CCCTGCGAGCTCTTCTGACTGCTCATTTATGGCAGTACTCGCGCGTGTCTGACCA 768
 Qy 241 TyrSerThrPheTyr 245
 Db 769 TACAGCACCTTCTAC 783

RESULT 2

BF503420 701 bp mRNA linear EST 19-APR-2001
 AT19406 Sprime AT Drosophila melanogaster adult testes pOTB7
 Drosophila melanogaster cDNA clone AT19406 5 similar to OsiStt3:
 Fnan0007748 'enzyme' located on: 3R 96B16-96B17; 04/09/2001, mRNA
 sequence.

ACCESSION

BF503420
 VERSION BF503420.2 GI:113697340

KEYWORDS

EST.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 701)
 Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,
 Berlan, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
 Farfan, D., Frise, B., George, R., Gonzalez, M., Guarin, H., Harris, N.,
 Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C. J., Nunoo, J.,
 Pacis, J., Paragge, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
 Lewis, S. E., Celniker, S. and Rubin, G. M.
 BGP/HMI AT Drosophila EST Project
 Unpublished (2000)

TITLE

JOURNAL

COMMENT

On Dec 6, 2000 this sequence version replaced gi:11596721.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit Genomic AB003750: arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1: 04/09/2001

Plate: AT.194 row: A column: 6

High quality sequence stop: 603.

FEATURES

source

1..701
 /organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="AT19406"

/sex="male"

/dev_stage="0-3 day old Ore-R males"

/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
 AT.121-AT.319: DHS-alpha Tona"

/clone_lib="AT Drosophila melanogaster adult testes pOTB7"

/note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
 Site_2: XhoI; The mRNA for the testis library was made
 from testes and seminal vesicles hand dissected from 0-3
 day old Ore-R males. RNA kindly provided by the lab of
 Margaret Fuller. Sized fractionated cDNAs were directly
 ligated into pOTB7. Plasmid cDNA library."

ORIGIN

Alignment Scores:
 Pred. No.: 1,758-226 Length: 701
 Score: 221.00 Matches: 221
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 28.55% Indels: 0
 DB: 2 Gaps: 0

US-10-028-384-8 (1-774) x BF503420 (1-701)

QY 224 ValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrsTyrSerThr 243
 Db 2 GTCTTCGTACTCTCTATATGGCAGGTACTCGCGGGTCTGTGACCACTACAGCACC 61
 QY 244 PheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGlnProIle 263
 Db 62 TTTCTACATCTGGGACTGCTGTCTCCATGAGATCCCTTCGGGGATTCACCAACGATA 121
 QY 264 ArgThrSerGlnHisMetAlaLeuGlyValPheValLeuLeuMetAlaValAlaThr 283
 Db 122 CGCACCACTGAACATGCTGCGTGGGAGTGTGTGCTCTTATGCGCGTGGCCACC 181
 QY 284 LeuArgHisLeuGlnSerValLeuSerArgAsnGlnPheArgLysLeuPheIleValGly 303
 Db 182 TTGCGCCATTTGCATGCTGCTGCGCAACAGATTCGGAAGCTGTTCATCGTCGGC 241
 QY 304 GlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMetLeuGlyVal 323
 Db 242 GGATTGCTGCTGGGCTTGGGCTCTTTGTGCGCGTCTGCTGCTCACCATGCTGGCGTT 301
 QY 324 ValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHis 343
 Db 302 GTGGCCCGGTGGAGTGGAGCTTCTACCTGCTGTGGATCTGCTACGCAAGATCCAC 361
 QY 344 IleProIleAlaSerValSerGlnHisGlnProThrThrTrpPheSerPhePhe 363
 Db 362 ATTCCCATCAATGCATCGCTGCGAGCATCAGCCACCACCTTGTTCTGTTCTCTTT 421
 QY 364 AspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIleLysGlnIle 383
 Db 422 GATCTGCACATCTGCTGGTGGCTTCCAGTGGAGTGTGCTACTGCATCAAGCAATC 481
 QY 384 AsnAspGluArgValPheValValLeuTyrAlaLysSerAlaValTyrPheAlaGlyVal 403
 Db 482 AACGACGAGCGCGTTTCTGCTGCTGTACGCCATCAGTGGCGTTTACTTCGCTGGTGTG 541
 QY 404 MetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPhe 423
 Db 542 ATGGTGGTGTGATGTTGACCTCAGCGCGTGTGTGTCATGCTGGCGGAGTGGCGTTT 601
 QY 424 SerGlyLeuLeuAspValPheLeuGlnGluAspSerSertysArgMetGlyThrAlaIle 443
 Db 602 TCGGACATCTGGATGTCTCTGCAAGAGGATTCGTTAAGCAATGGGCACAGCCATA 661
 QY 444 Ser 444
 Db 662 AGC 664

RESULT 3

BF496296

LOCUS

DEFINITION

655 bp mRNA linear EST 19-APR-2001
 AT10060, Sprime AT Drosophila melanogaster adult testes pOTB7
 Drosophila melanogaster cDNA clone AT10060 5' similar to OstStt3:
 FBan007748 'enzyme' located on: 3R 96B16-96B17; 04/08/2001, mRNA
 sequence.

ACCESSION

BF496296

VERSION

BF496296.2

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 655)

Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,

Bernan, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,

Farfan, D., Frise, E., George, R., Gonzalez, M., Guatin, H., Harris, N.,

Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J.,

Pacib, J., Paragosa, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,

Lewis, S.E., Ceiniker, S. and Rubin, G.M.

TITLE

JOURNAL

COMMENT

BDGP/HMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11579597.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AEO03750: arm3R [20671983, 20899380]

estimated-cyto:96B10-96C1: 04/08/2001 hit P element 1(3)j2D9:

1(3)j2D9 AQ026308 inserted at base 292 5', end of P element Inverse

PCR: 02/09/2001

Plate: AT 100 row: E column: 12

High quality sequence stop: 577.

FEATURES

source

1..655

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="AT10060"

/sex="male"

/dev_stage="0-3 day old Ore-R males"

/lab_host="Plates AT 10-AT 120: DH5-alpha. Plates

AT 121-AT 319: DH5-alpha 10x"

/note="Organ: ADULT testes; Vector: pOTB7; Site: 1: EcoRI;

Site 2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into pOTB7. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

Db:

Length:

Matches:

Conservative:

Mismatch:

Indels:

Gaps:

5.59e-219

214.00

100.00%

100.00%

27.65%

2

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0


```

Db      434 TCCGGGGCGCGCGCTTTCGCCCGCAGCTTCATGCCATGCTGCTGGCTACATCAGT 493
Qy      161 ArgSerValAlaGlySerTyAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      494 AGGTCGGTGGCTGGATCGTACGATAACGAGGCGATTGCCATATTCGCCCTGCAGTTCACC 553
Qy      181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheThrSerAlaAlaAla 200
Db      554 TACTTCCTGGTGGGTCGCTAGTGAACAGCTGGATCCGTTTCGGTGGCGCCGACCGCT 613
Qy      201 LeuSerTyPheTyMetValSerAlaTrpGlyGlyTyTrVal 214
Db      614 TTATCTACTTCTCATGATGGTGGTCCGCTGGGCTGCTAGCTG 655

RESULT 4
LOCUS   CK657451 715 bp mRNA linear EST 30-JAN-2004
DEFINITION Drosophila melanogaster larval-early pupal pOT2
ACCESSION CK657451
VERSION   CK657451.1 GI:41400976
KEYWORDS EST.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 715)
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
UNPUBLISHED (2001)
CONTACT: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: LP.237 row: E column: 11
High quality sequence stop: 644.
Location/Qualifiers
1..715
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP23759"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcORI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

```

FEATURES

```

source
1..715
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP23759"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DHS-alpha"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/note="Organ: whole body; Vector: pOT2; Site 1: EcORI;
Site 2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

ORIGIN
Alignment Scores:
Pred. No.: 6,18-219 Length: 715
Score: 214.00 Matches: 214
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.65% Indels: 0
DBs: 7 Gaps: 0

US-10-028-384-8 (1-774) x CK657451 (1-715)

```

```

Qy      1 MethanArgThrProLysMetLeuAsnSerIysValAlaGlyTySerSerLeulleThr 20
Db      60 ATGAATCGGACCGCCGAGATGCTGAACAGCAAGCTGGCTGCTACAGCAGCCTAATCACC 119
Qy      21 PheAlaIleLeuLeulleAlaIleTrpLeuAlaGlyPheSerSerArgLeuPheAlaIle 40

```

```

Db      120 TTCCGCATCTCTCTAATGCTGCTGGCTGGCCGGAATTTCTCTCCGCTCTTTCGGCGTCATC 179
Qy      41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyArgAlaThrAlaTy 60
Db      180 CGTTTCGAGTCGATATTCATGAGTTTCATCCGTGGTTCAACTACCGGGGCGACCGGCTAC 239
Qy      61 MetValGlnAsnGlyTyTrpTyAsnPheLeuAsnTrpPheAspGluArgAlaIleTrpTyPro 80
Db      240 ATGGTCGAGAAATGGTTGGTACAACTTCCTCACTGGTTCGACGAGCGCGGATGGTATCCG 299
Qy      81 LeuGlyArgIleValGlyGlyThrValTyTrpGlyLeuMetIleThrSerGlyGlyIle 100
Db      300 CTCCGCGAGGATTCGCGCGGTACCGTCTATCCGCGCTGATGATTAGCTCCGCGGGAATC 359
Qy      101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db      360 CATTGGCTCTCTCAAGTACTCAACATACCGGTCCATATTCTGTGACATCTGGGTGTTCTG 419
Qy      121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyTrpLeuLeuThrLysGluLeuTrp 140
Db      420 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 479
Qy      141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyTrpIleSer 160
Db      480 TCCGCGGGCGCGCGCTCTTTCGCCGCGACCTTCATCCGCCATCGTGCTGCTGCTACATCAGT 539
Qy      161 ArgSerValAlaGlySerTyAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      540 AGGTGCGTGGCTGGATCGTACGATAACGAGGCGCATTCGCATATTTCGCCCTTCAGTTCCACC 599
Qy      181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAla 200
Db      600 TACTTCTCTGCTGGCTGCTCAGTGAAGACTGATCCGCTGTTCTGTGCTGGCGCGGAGCGCT 659

RESULT 5
LOCUS   CK657649 679 bp mRNA linear EST 30-JAN-2004
DEFINITION LP24119.5prime LP Drosophila melanogaster larval-early pupal pOT2
ACCESSION CK657649
VERSION   CK657649
KEYWORDS EST.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 679)
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
UNPUBLISHED (2001)
CONTACT: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: LP.241 row: B column: 7
High quality sequence stop: 670.
Location/Qualifiers
1..679
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="LP24119"
/sex="male and female"
/dev_stage="larvae-pupae"

```

| | | | | | | | | | | | | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| QY | 1 | Met | Asn | Arg | Thr | Pro | Leu | Ser | Leu | Asn | Ser | Leu | Ser | Val | Ala | Gly | Tyr | Ser | Ser | Leu | Leu | Thr | 20 |
| DB | 56 | AT | GA | T | CGG | AGC | CGC | GA | AG | T | CCT | GA | A | C | AG | A | G | T | G | G | T | C | 115 |
| QY | 21 | Phe | Ala | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | Leu | 40 |
| DB | 116 | TT | CG | CA | T | C | T | T | G | C | T | T | G | C | T | T | G | C | T | T | G | C | 175 |
| QY | 41 | Arg | Phe | Glu | Ser | Leu | Leu | His | Glu | Phe | Asp | Pro | Tyr | Phe | Asn | Tyr | Arg | Ala | Thr | Ala | Tyr | 60 | |
| DB | 176 | CG | TT | CG | AG | T | CG | AT | T | CC | A | G | T | T | T | G | A | T | T | CA | A | T | 235 |
| QY | 61 | Met | Val | Gln | Asn | Gly | T | Trp | Tyr | Asn | Phe | Leu | Asn | Trp | Phe | Asp | Glu | Arg | Ala | Trp | Tyr | Pro | 80 |
| DB | 236 | AT | GG | T | GC | A | GA | T | G | T | T | G | T | T | G | T | T | C | A | A | C | T | 295 |
| QY | 81 | Leu | Gly | Arg | Leu | Val | Gly | Gly | Thr | Val | Tyr | Pro | Gly | Leu | Met | Leu | Thr | Ser | Gly | Gly | Leu | 100 | |
| DB | 296 | CT | CG | C | AG | AG | T | T | G | G | C | G | G | T | A | C | CG | T | GA | T | T | AC | 355 |
| QY | 101 | His | Trp | Leu | Leu | His | Val | Leu | Asn | Leu | Pro | Val | His | Leu | Arg | Asp | Leu | Cys | Val | Phe | Leu | 120 | |
| DB | 356 | CAT | T | GG | T | G | T | G | C | AC | G | T | CA | A | C | AC | CG | T | CA | T | T | CG | 415 |
| QY | 121 | Ala | Pro | Leu | Phe | Ser | Gly | Leu | Thr | Ser | Leu | Ser | Thr | Tyr | Leu | Leu | Thr | Leu | Ser | Glu | Leu | 140 | |
| DB | 416 | GC | GC | CA | T | CT | CA | G | T | GC | CT | GA | C | T | CC | A | T | CT | CC | AC | T | CT | 475 |
| QY | 141 | Ser | Ala | Gly | Ala | Gly | Leu | Phe | Ala | Leu | Ser | Phe | Leu | Ala | Leu | Val | Pro | Gly | Tyr | Leu | Ser | 160 | |

| | | | |
|--|---|--|-----|
| 476 | | TCGCGGCGCGCGCTCTTCGCGCGCGAGCTTCATCGCCATCGTCTGCTGCTACATCAGT | 535 |
| DB | | | |
| US-10-028-384-8 (1-774) x BI171940 (1-652) | | | |
| QY | 161 | ArgSerValAlaGlySerTyrAspAsnGluGlyLeuAlaLeuPheThr | 180 |
| DB | 536 | AGGTCGGTGGTGGATCGTACGATACAGGAGGATTCGCAATTCGCCCTGAGTTCACC | 595 |
| QY | 181 | TyrPheLeuTrpValArgSerVallyThrGlySerValPheTrpSerAlaAlaAla | 200 |
| DB | 596 | TACTTCCTGTCGGTGGCTCAGTGAAGACTGGATCCGTGTTCGTGCGCGCGAGCGCT | 655 |
| QY | 201 | LeuSerTyrPheTyrMetValSer | 208 |
| DB | 656 | TGTGCTACTTCTACATGGTGTCC | 679 |
| RESULT 7 | | | |
| LOCUS | BI171940 | | |
| DEFINITION | BI171940 652 bp mRNA linear EST 09-JUL-2001 | | |
| ACCESSION | RE13841.5 Prime RE Drosophila melanogaster normalized Embryo pP1c-1 | | |
| VERSION | Drosophila melanogaster cDNA clone RE13841.5 similar to OstStt3: | | |
| KEYWORDS | Fban007748 'enzyme' located on: 3R 96B16-96B17; 04/06/2001, mRNA | | |
| ORGANISM | sequence. | | |
| SOURCE | BI171940 | | |
| ORGANISM | BI171940.1 GI:14637747 | | |
| SOURCE | EST. | | |
| ORGANISM | Drosophila melanogaster (fruit fly) | | |
| SOURCE | Drosophila melanogaster | | |
| ORGANISM | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | |
| SOURCE | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| ORGANISM | Ephydroidea; Drosophilidae; Drosophila. | | |
| SOURCE | 1 (bases 1 to 652) | | |
| REFERENCE | Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., | | |
| AUTHORS | Carlson, J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., | | |
| | George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., | | |
| | Misra, S., Mungall, C. J., Nuroo, J., Pacleb, J., Paragas, V., Park, S., | | |
| | Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and | | |
| | Rubin, G. M. | | |
| TITLE | BDGP/HMI RE Drosophila EST Project | | |
| JOURNAL | Unpublished (2001) | | |
| COMMENT | Contact: Stapleton, M. | | |
| BDGP | | | |
| Lawrence Berkeley National Lab | | | |
| One Cyclotron Rd, Berkeley, CA 94720, USA | | | |
| Fax: 510 486 6798 | | | |
| Email: http://www.fruitfly.org/EST_estofruitfly.berkeley.edu | | | |
| hit genomic A8003750; arm:3R [20671983.20899380] | | | |
| estimated-cyto:96B10-96B1: 04/06/2001 hit P element 1(3)j2D9: | | | |
| 1(3)j2D9 A9026308 inserted at base 232 5', end of P element Inverse | | | |
| PCR: 04/06/2001 | | | |
| Plate: RE.138 row: D column: 5 | | | |
| High quality sequence stop: 590. | | | |
| Location/Qualifiers | | | |
| 1. 652 | | | |
| /organism="Drosophila melanogaster" | | | |
| /mol_type="mRNA" | | | |
| /db_xref="taxon:7227" | | | |
| /clone="RE13841" | | | |
| /sex="male and female" | | | |
| /dev_stage="0-24 hours mixed stage embryonic" | | | |
| /lab_host="DH5-alpha TonA" | | | |
| /clone_lib="RE Drosophila melanogaster normalized Embryo | | | |
| pP1c-1" | | | |
| /notes="Organ: embryo; Vector: pP1c1; Site: 1: XhoI; Site 2: | | | |
| BamHI; Library was kindly generated by Piero Carninci at | | | |
| the RIKEN. The library was normalized and excised using | | | |
| Cre recombinase. Plasmid cDNA library." | | | |
| ORIGIN | | | |
| Alignment Scores: | | | |
| Pred. No.: 1.11e-200 Length: 652 | | | |
| Score: 197.00 Matches: 197 | | | |
| Percent Similarity: 100.00% Conservative: 0 | | | |
| Best Local Similarity: 100.00% Mismatches: 0 | | | |
| Query Match: 25.45% Indels: 0 | | | |

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit genomic AB003750: arm:3R [20671983,20899380]
 estimated-cyto:96B10-96C1: 05/13/2001 hit P element 1(3)J2D9;
 1(3)J2D9 A0026308 inserted at base 292 5' end of P element inverse
 PCR: 05/13/2001
 Plate: RE.434 row: C column: 1
 High quality sequence stop: 563.
 Location/Qualifiers

FEATURES

source

1. .666

/organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE43425"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:
 Pred. No.: 1,358-199 Length: 666
 Score: 196.00 Matches: 196
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.32% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x B1357074 (1-666)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
 Db 79 ATGAATCGGACCGCGAGATCTGAACAGCAAGTGGCTGCTACAGCCTTAATCACC 138
 Qy 21 PheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValle 40
 Db 139 TTCGCCATCTGCTAAATCGCGCTGGCTGGCGGATTTTCCTCTCGCTCTTCGCGCTCATC 198
 Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 Db 199 CGTTTCGAGTGCATTATCCATGAGTTTGATCCGGTGGTTCACTACCGGGCCACCGCTAC 258
 Qy 61 MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
 Db 259 ATGTGTCAGAAATGGTTGGTACAACTTCTCAACTGGTTTCGACGAGCGCGCATGGTATCCG 318
 Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 319 CTCGGCAGGATTGGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGCGAATC 378
 Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 379 CATTTGGCTGTGTCAGCTACTCAACATACCGGTTCATATTCGTGACATCTGGGTTCCTG 438
 Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 439 GCGCCGATCTTCAGTGGCCCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 498
 Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 499 TCGCGGGCGCGCCCTTCCTTCGCGCGAGCTTCATGCGCATCTGCTGGCTACATCAT 558
 Qy 161 ArgSerValAlaGlySerTyrAspAsnGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 559 AGTCGGTGGCTGGATCGTACGATAACAGAGGCAATGCCATATTCGCCCTGCAGTTTACC 618
 Qy 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
 Db 619 TACTTCTGTGGGTGGCTCAGTGAAGACTGGATCGGTGTTCTGGTGG 666

RESULT 9

B1227902
 LOCUS

DEFINITION

B1227902 668 bp mRNA linear EST 11-JUL-2001
 RE25288.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
 Drosophila melanogaster cDNA clone RE25288 5 similar to OatStt3;
 FBan0007748 'enzyme' located on: 3R 96B16-96B17: 04/12/2001, mRNA
 sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 668)

Capleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,

Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,

George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,

Misra,S., Mungall,C.J., Nuno,J., Pacleb,J., Paragas,V., Park,S.,

Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and

Rubin,G.M.

BDGP/HMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AB003750: arm:3R [20671983,20899380]

estimated-cyto:96B10-96C1: 04/12/2001 hit P element 1(3)J2D9;

1(3)J2D9 A0026308 inserted at base 292 5' end of P element inverse

PCR: 03/20/2001 row: H column: 4

Plate: RE.252 row: H column: 4

High quality sequence stop: 520.

Location/Qualifiers

1. .668

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE25288"

/sex="male and female"

/dev_stage="0-24 hours mixed stage embryonic"

/lab_host="DHS-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 1,358-199 Length: 668

Score: 196.00 Matches: 196

Percent Similarity: 100.00% Conservativity: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 25.32% Indels: 0

DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x B1227902 (1-668)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20

Db 81 ATGAATCGGACCGCGAGATCTGAACAGCAAGTGGCTGCTACAGCCTTAATCACC 140

Qy 21 PheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValle 40

Db 141 TTCGCCATCTGCTAAATCGCGCTGGCTGGCGGATTTTCCTCTCGCTCTTCGCGCTCATC 200

Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60

| | |
|---|--|
| ORIGIN | |
| Alignment Scores: | Length: 641 |
| Pred. No.: | 1558-198 |
| Score: | 195.00 |
| Percent Similarity: | 100.00% |
| Best Local Similarity: | 100.00% |
| Query Match: | 25.13% |
| DB: | 1 |
| US-10-028-384-8 (1-774) x AI257750 (1-641) | |
| QY | 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20 |
| DB | 55 ATGAATCGGACGCGAAGATGCTGAACAGCAAGGTGGCTACAGCAGCTAATCACC 114 |
| QY | 21 PheAlaIleuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40 |
| DB | 115 TTCGCATCTGCTAATCGCCTGGCTGGCCGGAATTTCTCTCGCTCTTCGCCCTCATC 174 |
| QY | 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60 |
| DB | 175 CGTTTCGAGTCGATATATCCATGAGTTTGATCCGTGGTTCACTACCGGCGCACCGCTAC 234 |
| QY | 61 MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80 |
| DB | 235 ATGGTGCAGAAATGGTTGGTACAACTTCCCTCAACTGGTTCGACGCGCGCATGGTATCG 294 |
| QY | 81 LeuGlyArgIleValGlyGlyThrValTyrPcGlyLeuMetIleThrSerGlyGlyIle 100 |
| DB | 295 CTGGCAGGAGTTGGCGGCTGCTATCCGCGCTGATGATTACGTCCGCGGAATC 354 |
| QY | 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120 |
| DB | 355 CATTGGCTGTCGACGCTACTCAACATACCGGTCCATATTCGTGACATCTCGCGTTCCTG 414 |
| QY | 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140 |
| DB | 415 GCGCGGATCTTCAGTGGCTGACCTCATCTCCACTACTCTGCTGACCAAGAGCTGTGG 474 |
| QY | 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160 |
| DB | 475 TCCGCGGCGCGCGCTCTTCGCGCGCAGCTTTCATCGCATCTGCTGCTGGCTACATCAGT 534 |
| QY | 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180 |
| DB | 535 AGTTCGCTGGTGGATCGTACGATAACGAGGCGCATTCGCCATATTCGCCCTGCGATTAC 594 |
| QY | 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrp 195 |
| DB | 595 TACTTCTGTGGTGGCTCGCTCAGTGAAGACTGGATCCGTGTCTCG 639 |
| RESULT 11 | |
| BG641172 | |
| LOCUS | |
| DEFINITION | |
| SD12448.5 prime SD Drosophila melanogaster Schneider L2 cell culture | |
| POT2 Drosophila melanogaster cDNA clone SD12448 5 similar to | |
| OtsSt3: FBAN007748 'enzyme' located on: 3R 96B16-96B17; | |
| 04/13/2001, mRNA sequence. | |
| ACCESSION | |
| BG641172 | |
| VERSION | |
| BG641172.1 GI:13773098 | |
| KEYWORDS | |
| EST | |
| SOURCE | |
| ORGANISM | |
| Drosophila melanogaster (fruit fly) | |
| Drosophila melanogaster | |
| Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | |
| Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | |
| Ephydroidea; Drosophilidae; Drosophila. | |
| REFERENCE | |
| 1 (bases 1 to 663) | |
| Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., | |
| Lewis, S. and Rubin, G.M. | |
| BDGP/HMI Drosophila EST Project | |
| JOURNAL | |
| Unpublished (2001) | |
| COMMENT | |
| Contact: Stapleton, M. | |

| | |
|---|--|
| FEATURES | |
| Source | |
| 1..641 | |
| /organism="Drosophila melanogaster" | |
| /mol_type="mRNA" | |
| /db_xref="taxon:7227" | |
| /clone="LP06212" | |
| /sex="male and female" | |
| /dev_stage="larvae-pupae" | |
| /lab_host="DH5-alpha" | |
| /clone_lib="LP Drosophila melanogaster larval-early pupal | |
| POT2 | |
| /note="Organ: whole body; Vector: POT2; Site 1: EcoRI; | |
| Site 2: XhoI; Sized fractionated cDNAs were directly | |
| ligated into POT2. plasmid cDNA library. | |

| | |
|---|--|
| RESULT 10 | |
| AI257750 | |
| LOCUS | |
| DEFINITION | |
| LP06212.5 prime LP Drosophila melanogaster larval-early pupal POT2 | |
| Drosophila melanogaster cDNA clone LP06212 5prime, mRNA sequence. | |
| ACCESSION | |
| AI257750 | |
| VERSION | |
| AI257750.1 GI:3865275 | |
| KEYWORDS | |
| EST | |
| SOURCE | |
| Drosophila melanogaster (fruit fly) | |
| Drosophila melanogaster | |
| Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | |
| Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | |
| Ephydroidea; Drosophilidae; Drosophila. | |
| REFERENCE | |
| 1 (bases 1 to 641) | |
| Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., | |
| Lewis, S. and Rubin, G.M. | |
| BDGP/HMI Drosophila EST Project | |
| JOURNAL | |
| Unpublished (2001) | |
| COMMENT | |
| Contact: Stapleton, M. | |
| BDGP | |
| Lawrence Berkeley National Lab | |
| One Cyclotron Rd, Berkeley, CA 94720, USA | |
| Fax: 510 486 6798 | |
| Email: http://www.fruitfly.org/EST , est@fruitfly.berkeley.edu | |
| hit P element sequence 1(3)J2D9 | |
| Plate: 62 row: A column: 12 | |
| High quality sequence stop: 462. | |
| Location/Qualifiers | |
| 1..641 | |
| /organism="Drosophila melanogaster" | |
| /mol_type="mRNA" | |
| /db_xref="taxon:7227" | |
| /clone="LP06212" | |
| /sex="male and female" | |
| /dev_stage="larvae-pupae" | |
| /lab_host="DH5-alpha" | |
| /clone_lib="LP Drosophila melanogaster larval-early pupal | |
| POT2 | |
| /note="Organ: whole body; Vector: POT2; Site 1: EcoRI; | |
| Site 2: XhoI; Sized fractionated cDNAs were directly | |
| ligated into POT2. plasmid cDNA library. | |

| | |
|----|---|
| Db | 201 CGTTTCGAGTCGATATCATGAGTTTGATCCGTGGTTCAACTACCGGCGCACCGCTAC 260 |
| QY | 61 MetValGlnAsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80 |
| DB | 261 ATGGTGCAGAAATGGTTGGTACAACTTCTCAACTGTTTCGACGAGCGCGCATGGTATCCG 320 |
| QY | 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100 |
| DB | 321 CTCGCGCAGGATGTGGCGGCTACCGCTATATCCCGCCTGATGATACGTCGCGCGGAATC 380 |
| QY | 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120 |
| DB | 381 CATTGGCTGCTGCACGCTACTCAACATACCGGTCCATATTCGTGACATCTGGTGTCTCG 440 |
| QY | 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140 |
| DB | 441 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGTGG 500 |
| QY | 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160 |
| DB | 501 TCCGCGGCGCGCGCTCTTCGCGCGCAGCTTCATCGCCATCTGCTGCTGCTGCTACATCAGT 560 |
| QY | 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180 |
| DB | 561 AGTTCGCTGGTGGATCGTACGATAACGAGGCGATTGCCATATTCGCCCTGCGATTAC 620 |
| QY | 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196 |
| DB | 621 TACTTCTGTGGTGGCTCAGTGAAGACTGGATCGGTGTTCTGTGTCG 668 |

BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AE003750: arm:3R [20671983..20899380]

estimated-cyto:96B10-96C1: 04/13/2001 hit P element 1(3)J2D9:

1(3)J2D9 AQ026308 inserted at base 292 5' end of P element Inverse

PCR: 03/20/2001

Plate: SD.124 row: D column: 12

High quality sequence stop: 604.

Location/Qualifiers

FEATURES

Source

1..663

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clones="SD12448"

/lab_hosts="DHS-alpha"

/clone_lib="SD Drosophila melanogaster Schneider L2 cell

culture pot2"

/notes=vector: pot2; Site.1: EcoRI; Site 2: XhoI; Sized

fractionated cDNAs were directly ligated into pot2.

Plasmid cDNA library."

ORIGIN

Alignment Scores:

Pred. No.: 1.6e-198 Length: 663
Score: 195.00 Matches: 195
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.19% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x BG641172 (1-663)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerLeuLeuThr 20
Db 71 ATGAATCGGACGCGAGATGCTGACAGCAAGTGGTGGCTACAGCGCTTAATCACC 130
Qy 21 PheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db 131 TTCGCCATCTCGTAATCGCGCTGGCTGGCGGATTTTCCTCTCGCTCTTCGCGGTCATC 190
Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrAlaTyr 50
Db 191 CGTTTCAGTCGATATCCAGATTTTGAATCCGCTGTTCAACTACCGGGCCACCGCTAC 250
Qy 61 MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db 251 ATGGTGCAGAAATGGTGGTAACTTCCTCAACTGCTTCGACGAGCGCATGGTATCCG 310
Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 311 CTGGCAGGATTGTGGCGGTACCGTCTATCCCGGCTCATGATTACGTCGCGCGGAATC 370
Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 371 CATTTGGCTGTGCACGTACTCAACATACCGTTCATTCGTGACATCTGCTGTCTCTG 430
Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIysGluLeuTrp 140
Db 431 GCGCCGATCTTCAGTGGCGCTGACCTTCATCTCCACCTACCTGCTGACCAAGGAGCTGG 490
Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 491 TCCGGGGCGCGGCTCTTCGCGCGCAGCTTCATCCCATCTGCTGCTGCTGCTACATCAGT 550
Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 551 AGGTCGGTGGCTGGATCGTACGATACAGGAGGCAATGTCATATTCGCTTCAGTTTACC 610
Qy 181 TyrPheLeuTrpValArgSerValIysThrGlySerValPheTrp 195

Db

611 TACTTTCTGCTGGTGGCTCAGTGAAGACTGATCGGTGTTCTGG 655

RESULT 12

BF504378

LOCUS

DEFINITION

BF504378 616 bp mRNA linear EST 19-APR-2001
AT05729, Sprime AT Drosophila melanogaster adult testes potB7
Drosophila melanogaster cDNA clone AF05729 5 similar to Oststct3:
FBan0007748 'enzyme' located on: 3R 96B16-96B17;: 04/07/2001, mRNA
sequence.

ACCESSION

BF504378

VERSION

BF504378.2

KEYWORDS

EST.

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 616)

REFERENCE

AUTHORS

Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E.,

Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,

Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,

Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J.,

Paciel,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,

Lewis,S.E., Celniker,S. and Rubin,G.M.

BDGP/HEMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11587679.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic AE003750: arm:3R [20671983..20899380]

estimated-cyto:96B10-96C1: 04/07/2001 hit P element 1(3)J2D9:

1(3)J2D9 AQ026308 inserted at base 292 5' end of P element Inverse

PCR: 02/09/2001

Plate: AT.57 row: C column: 5

High quality sequence stop: 592.

Location/Qualifiers

1..616

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clones="AT05729"

/sex="male"

/dev_stages="0-3 day old Ore-R males"

/lab_hosts="Plates AT.10-AT.120; DHS-alpha. Plates

AT.121-AT.319; DHS-alpha Tona"

/clone_lib="AT Drosophila melanogaster adult testes potB7"

/notes=Organ: ADULT testes; Vector: potB7; Site.1: EcoRI;

Site.2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into potB7. Plasmid cDNA library."

ORIGIN

Alignment Scores:
Pred. No.: 1.77e-197 Length: 616
Score: 194.00 Matches: 194
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.06% Indels: 0
DB: 2 Gaps: 0

US-10-028-384-8 (1-774) x BF504378 (1-616)

Qy 2 AsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThrPhe 21
Db 33 AATCGGACGCGAGATGCTGACAGCAAGTGGCTGCTGCTACAGCGCTTAATCAGCTTC 92
Qy 22 AlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArg 41

Db 93 GCCATCTCTGTAATCGCTGCTGGCGGATTTTCCTCTCGCCCTTTTCGGCGGCAATCCGT 152
 Qy 42 PheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMet 61
 Db 153 TTCGAGTCGATATTCATGAGTTTGATCCGTTGGTTCACTACCGGCGCACCGCTACATG 212
 Qy 62 ValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrProLeu 81
 Db 213 GTGAGAAATGTTGTGTACAACTTCCTCAACTGGTTCGACGAGCGCGATGTATCCGCTC 272
 Qy 82 GlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHis 101
 Db 273 GGCAGATTTGGGCGGTACCGTCTATCCCGCCTGATGATTAGCTCGGCGGATTCAT 332
 Qy 102 TrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeuAla 121
 Db 333 TGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTCTCTGGCG 392
 Qy 122 ProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyrSer 141
 Db 393 CCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGCTGCTGCTGCTGCTGCTG 452
 Qy 142 AlaGlyAlaGlyLeuPheAlaIleSerPheIleAlaIleValProGlyTyrIleSerArg 161
 Db 453 GCGGGCGCGCGCTCTTCGCGCGCAGCTTCATCGGCATCTGCTGCTGCTGCTGCTGCTG 512
 Qy 162 SerValAlaGlySerTyrAsnAsnGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
 Db 513 TCGTGTGCTGATCGTACGATACGAGGCGCATTCGCATATTCGCCTCGATTCACCTAC 572
 Qy 182 PheLeuTyrValArgSerValLysThrGlySerValPheTyr 195
 Db 573 TTCTCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 614

RESULT 13
 AI295381 630 bp mRNA linear EST 19-APR-2001
 LOCUS LP08987.5prime LP Drosophila melanogaster larval-early pupal pot2
 DEFINITION Drosophila melanogaster cDNA clone LP08987 5prime, mRNA sequence.

ACCESSION AI295381.1 GI:3944788
 VERSION EST.
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 630)
 AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.
 TITLE BDGP/HMMI Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 hit p element sequence 1(3)2D9
 Plate: 89 row: H column: 3
 High quality sequence stop: 498.

FEATURES
 source
 1. 630
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="LP08987"
 /sex="male and female"
 /dev_stage="larvae-pupae"
 /lab_host="DHS-alpha"
 /clone_lib="LP Drosophila melanogaster larval-early pupal pot2"

/note="Organ: whole body; Vector: pot2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."

ORIGIN

Alignment Scores:

Prod. No.: 3.07e-194 Length: 630
 Score: 191.00 Matches: 191
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 24.68% Indels: 0
 DB: 1 Gaps: 0

US-10-028-384-8 (1-774) x AI295381 (1-630)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 Db 56 ATGAATCGGACCGCCGAGATGCTGAACGAGGTTGGCTGCTACGAGGCTATATCACC 115
 Qy 21 PheAlaIleLeuLeuIleAlaTyrPheAlaGlyPheSerSerArgLeuPheAlaValIle 40
 Db 116 TTCGCATCTCTGTAATCGCTGGCTGGCGGATTTCTCTCGGCTCTTCGCCGTATC 175
 Qy 41 ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
 Db 176 CGTTTCGAGTCGATATTCATGAGTTGATCGGTGTTCACTACCGGCGCACCCCTAC 235
 Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80
 Db 236 ATGCTGCAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 295
 Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 296 CTCGGCAGGATTTGTCGGCGGTACCGCTCTATCCCGGCTGATGATTACGTCGGCGGAATC 355
 Qy 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 356 CATGTGCTGCTGCACGTACTCAACATACCGTCCATATTCGTGACATCTGCGTGTCTCG 415
 Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
 Db 416 GCGCGCATCTTCAGTCGCTGACCTCCATCTCCACCTACCTGCTGCTGCTGCTGCTG 475
 Qy 141 SerAlaGlyAlaGlyLeuPheAlaIleSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 476 TCCGCGGCGCGCGCTCTTCGCGCGCAGCTTCATGCCATCGCTGCTGCTGCTGCTG 535
 Qy 161 ArgSerValAlaGlySerTyrAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 536 AGTCCGTGCTGATCGTACGATACGAGGCGCATTCGCATATTCGCCCTGCAATTCACC 595
 Qy 181 TyrPheLeuTyrValArgSerValLysThrGly 191
 Db 596 TACTTCTGTGGTGGCTCAGTGAAGACTGGA 628

RESULT 14

BG636414

LOCUS

DEFINITION

SD14123.5prime SD Drosophila melanogaster Schneider L2 cell culture

POT2 Drosophila melanogaster cDNA clone SD14123 5 similar to

OstStt3: FBan0007748 'enzyme' located on: 3R 96B16-96B17;

04/13/2001, mRNA sequence.

ACCESSION BG636414

VERSION BG636414.1

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 644)

AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

BG636414 644 bp mRNA linear EST 23-APR-2001
 SD14123.5prime SD Drosophila melanogaster Schneider L2 cell culture
 POT2 Drosophila melanogaster cDNA clone SD14123 5 similar to
 OstStt3: FBan0007748 'enzyme' located on: 3R 96B16-96B17;

04/13/2001, mRNA sequence.

ACCESSION BG636414

VERSION BG636414.1

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 644)

AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,

1 MetAsnArgThrProLysMetIleuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
81 ATGAAATCGGACCCGACATGTTGACACAGCAGGTGGTGGCTACAGCAGCTAATCACC 140

Thu Dec 16 16:25:15 2004

```

QY      21  PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
Db      |||
Db      141  TTGGCCATCTGCTAAATCGCTGGCTGGCGGATTTTCCTCTCGCCTCTTCGCGGTATC 200
QY      41  ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
Db      |||
Db      201  CGTTTCGAGTCGATTATCATGAGTTTGATCCGGTTCAACTACCGGGCCACCGCCTAC 260
QY      61  MetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db      |||
Db      261  ATGGTGCAGATGGTTGGTACAACTTCCTCAACTGGTTCGACGAGCGCATGGTATCCG 320
QY      81  LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db      |||
Db      321  CTCCGCGAGGATGTGGGGGTACCGTCTATCCCGCCTGATGATTACGTCGGCGGAATC 380
QY      101  HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db      |||
Db      381  CATTGGCTCTGCAGGTACTCAACATACGGTCCATATTCGTGCATCTGGTGTCTCTG 440
QY      121  AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIysGluLeuTrp 140
Db      |||
Db      441  GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGGAGCTGTGG 500
QY      141  SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db      |||
Db      501  TCCGGGGCGCGGCTCTTCGCGCCGAGCTTCATCGCCATCGTGCCTGGCTACATCAGT 560
QY      161  ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db      |||
Db      561  AGTCCGTGGCTGGATCGTACGATAACGAGGCAATGCCATATTCGCTTCAGTTCACC 620
QY      181  TyrPheLeuTrpValArgSerValIysThrGly 191
Db      |||
Db      621  TACTTCTGTGGTGGCTCAGTGAAGACTGGA 653

```

Search completed: December 15, 2004, 06:14:19
Job time : 6000.52 secs

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Lawrence Berkeley National Laboratory, One Cyclotron Road,

Berkeley, CA 94720, USA

Berkeley, CA 94720, USA
Sequence submitted by:

sequence submitted by:
Herkelev Drosophilae Genome project

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory

Lawrence Berkeley
Berkeley, CA 94720

Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from *Drosophila* Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

```

source
1..2417
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon.7227"

```

```
gene
<1. .2417
/locus_tag="taxon:1227"
/gene="OstStt3"
/db_xref="E1V3A5F.EB00011336"
```

```
CDS
/cd_xref=FLYBASE:FBNR0011336
<1..2395
/gene="OstStt3"
/note="Longest ORF"
/codon_start=2
/product="GMO1838p"
/protein_id="AAD27851.2"
```

```

/dec 1
/adb_xrszf="GI-28317200"
/translation="LSECEVPAQIGYFNPRVTGMWNMNRPKMLNKSVAGYSSLI/
AILLALMGFSRFLAVRFSIFIEHEDPMFNRYATAYQWNGVYFNLWNEDEAWY
PLGIRVIGTYPGLMITSQIWHLLHVLHNPVPHRIDCFLVAFIQLGSLITSYLLT
ELWAGAGLFAAAGPYFIAIPYFIVFVSVSYONEGATIPALQFYFLWARSVKTYGFW
SAAALAGYFVWGAWGYYFINLPLHGVYVILMGRYSYLTXYSTFYVILCLGFSM
OIPFGVQPIPTSEHMAALGVFLMVAATLHERQSVLSRSEFEKULFIVGGLLVGV
EVAVILTMGLVPAWGBSPYSLMDTGYAKHILPIASVSEHQPTWTFSEFDHLIY
CAPFVGWYCIKQINERVFVILIASVIFAGVWYRMLMLTIPVCMCLAGVAFSGLL
LVFQESXRMGMLTASATEVEDEASLEKTLVDKAGLKHRTKHDQDQDQGVSEEN
LKSVILAVLLMLMTFAVETPTNSVSSPIVLAFHNSQDSRNILDDFEARYWL
SCNTADBARVWNSYGYIGDGMARWTLVNDTNWNSHIALVKAAMSSTEBSKYEIM
TSLVDVYLVIFGVGYIGDGNKFLMWVRIAGEBHPKOIKESDYPTDQRFVDAE
GAPALLCMKYSLYRFBELKGLVGRSGYDTRNAVIGNKFDLTLYEEAAVTEHFW
LAVLVYVKKKHEFNBSIKTKSEPTIPANFISLRNKLKPGYFEPVDFWAKCPKTL"

```

ORIGIN

| Alignment Scores: | | | |
|--|-----------|---|------|
| Pred. No.: | 3.14e-314 | Length: | 2417 |
| Score: | 4046.00 | Matches: | 774 |
| Conservative: | 100.00% | Mismatches: | 0 |
| Best Local Similarity: | 100.00% | Indels: | 0 |
| Query Match: | 100.00% | Gaps: | 0 |
| DB: | 3 | | |
| JUS-10-028-384-8 (1-774) x AF132552 (1-2417) | | | |
| QY | 1 | MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr | 20 |
| DB | 71 | ATGAAATCGAGCGCCGAGATGCTGACACAGAGTGGCTGGCTACAGCAGCCATATACCC | 130 |
| QY | 21 | PheAlaIleLeuIleAlaThrLeuAlaGlyPheSerSerArgLeuPheAlaValIle | 40 |
| DB | 131 | TTCCCAATCTGTCTAAATCGCGTGGCTGGCGCGAATTTCTCTCGCCCTCTCCGCCGTCATC | 190 |
| QY | 41 | ArgPheGluSerIleIleHisGluPheAspProTTPheAsnTyrArgAlaThrAlaTyr | 60 |

QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 1331 GTGGCTTTTCGGGACTGTGGATGTTCTCTGCAAGAGGATTCGTCTAAGCGAATGGC 1390
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLeuLysThr 460
 DB 1391 ACAGCCATAGCGCGCCACCGAAGTGGATGAGCTGAGGATTCATTTGAGAGAGACG 1450
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 DB 1451 CTGTACGACAAAGCTGGCAAGCTGAAGCATGCTACTAAGCATGATGCCAGCAGGACT 1510
 QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValIleMetLeuLeuMetMet 500
 DB 1511 GGCGTCAGCTCAACCTGAGAGATATGTTATTTTGGCCGTTCTTAATGCTTGTATGATG 1570
 QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 DB 1571 TTGCGTGTCTCACTGACAGTGGGTGACCAAGCATGCTACTCCAGTCCCTCCATTTGCTTG 1630
 QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr 540
 DB 1631 GCTTTCCACAAAGCAAGATGATCCGCAACATTTTAGACGATTTCCAGAGGCTTAC 1690
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly 560
 DB 1691 TACTGCTTTTCGACAGAACACTCCGATGATGCTCGCTTATGTTCTTGGTGGATTTACGGA 1750
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 DB 1751 TACCAGATAGCGGGAATGCAACAGACAGACGCTAGTGGATATATATACGTGGACAT 1810
 QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 DB 1811 AGTCACATAGCGCTGTTGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAAT 1870
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 DB 1871 ATGACATCTTTGACGTGACATGCTTTTGGTGATCTTTGGCGGTGATCGGCTATCT 1930
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 DB 1931 GSCGATGATCAACAAAGTCTCTGTGGATGGTCCGAATTTGCTGAGGAGAGCATCCCAAG 1990
 QY 641 AsnIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
 DB 1991 GACATTAAGGAAGCGATTTACTTTTACCAGCCGCGTGAAATTCAGGGTAGATGCCGAAGT 2050
 QY 661 AlaProAlaLeuLeuAsnLysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 DB 2051 GCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGAATG 2110
 QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 DB 2111 AAGTTGGACTCAGAGGTCATCTGGATGATGATGCACACGTAACCGCGCTCAITGGGAAT 2170
 QY 701 LysAspPheAspLeuThrTyrLeuGluAlaTyrThrThrGluHisTrpLeuValArg 720
 DB 2171 AAGGACTTCGATCTGACCTACCTGGAGGAGGCTACACACAGAACTCTGGCTTTGTTGCG 2230
 QY 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 DB 2231 ATCTATAGGTGAAGAGCGGATGATTTCAATAGACCATCACTGAAGACCAAGAGAGA 2290
 QY 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760
 DB 2291 ACGATTCTCCAGCAAACTTCATTTCCGAAAGAACTCTAAGCGTCCGCAAGGGCTACATA 2350
 QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 DB 2351 CGAAACCGACCGGCTGTTGTTTAAAGGAAACCAACCTTGAAA 2392

RESULT 2
 AX799088
 LOCUS 2417 bp DNA linear PAT 08-OCT-2003
 DEFINITION Sequence 7 from Patent WO03054008.
 ACCESSION AX799088
 VERSION AX799088.1 GI:37605061
 KEYWORDS Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1
 AUTHORS Perreault, C. and McBride, K.
 TITLE Mammalian SAMP protein, gene sequence and uses thereof in cancer therapy
 JOURNAL Patent: WO 03054008-A 7 03-JUL-2003;
 COMPATIGENE Inc. (CA)
 FEATURES
 source
 1. 2417
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /note="S113, Acc# AF132552"
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,14e-314 Length: 2417
 Score: 4046.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-8 (1-774) x AX799088 (1-2417)
 QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 DB 71 ATGAAATCGAGCGCGAAGATGCTGAACAGCAAGGTGGTGGTACACAGCCTAATCACC 130
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 131 TTTCGCCATCTCTGTAATCGCTGGTGGCGGATTTTCCTCTCGCCTCTTCGCGGTATC 190
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 DB 191 CGTTTCGAGTCGATATCCATGAGTTGATCCGTTGTTCAACTACCGGGCCACCGCTAC 250
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80
 DB 251 ATGGTGCAGAAATGGTTGTACAACTTCCTCAACTGGTTTCGACGAGCGCATGGTATCCG 310
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 311 CTCGCGAGATTTGGCGGTACCGTCTATCCCGGCTGATGATTAGTTCGCGGATC 370
 QY 101 HistTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 371 CATTCGCTGCTGACGTACTCAACATACCGGTCCATATTCGTGACATCTCGCTGCTCG 430
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 DB 431 GGCGCGATCTTCAGTGGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGGAGCTGG 490
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 DB 491 TCCGCGGCGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 550
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 551 AGTTCGGTGGCTGGATCTGATACAGAGAGGCAATTCGCATATTCGCTCGCGGCTCAC 610
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheThrPheAlaAlaAla 200

| | | | | |
|----|------|---|--|-----|
| | 611 | TAC | TTCCTCGTGGGTGCGCCTCAGTGAAAGACTGGATCCGTGTCTCTGTGTCGGCCGCAGACGCCGT | 670 |
| Db | | | | |
| Qy | 201 | LeuSer | TyrPheYrMetValSerAlaIatrpIGlyTYrValPheillelleAsnLeuille | 220 |
| Db | 671 | TGTGCTCACTTCTACATGGTGTCCGCTCGGGGTGGCTACGTGTTTCATCATCAACCTGATA | 730 | |
| Qy | 221 | ProLeuHisValPheValLeuLeuilleMetGlyArgTYrSerProArgLeuLeuThrSer | 240 | |
| Db | 731 | CCCCGTGCAGCTCTTCGTACTCTCATTAATGGCAGAGTACTCCGCCGCTCTCTGTACCACGCG | 790 | |
| Qy | 241 | TyrSerThrPheTYrIlleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe | 260 | |
| Db | 791 | TACAGCACCTTCTACATCCTCGGACGTCTGTCTTCATCATGAGATCCCCTCGTGGGATTTC | 850 | |
| Qy | 261 | GlnProfileArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla | 280 | |
| Db | 851 | CAACCGATACGACCACTGAACATGGCTCGCTGGGNGTCTTTGTCTCTCTATCGCC | 910 | |
| Qy | 281 | VallalathrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLYLeuPhe | 300 | |
| Db | 911 | GTGGCCACCTTTCGCCCAATTTGCAGTCCGTGCTCGCGCAACGAGTTCCGGAGAGCTGTT | 970 | |
| Qy | 301 | IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet | 320 | |
| Db | 971 | ATCGTCGGCGGATGCTGTGTGGCGCTTGGCTCTTTGTGGCGCTGTGGTGTCTCACCATG | 1030 | |
| Qy | 321 | LeuGlyValValAlaProTrpSerGlyArgPheTYrSerLeuTroAspThrGlyTYrAla | 340 | |
| Db | 1031 | CTGGCGCTTGTGGCCCCGTGAGTAGACGCTTCTACTGCTGTGGGATCTGGGTACGCC | 1090 | |
| Qy | 341 | LysIleHisIleProfilelleAlaSerValSerGluHisGlnProThrThrtppPheSer | 360 | |
| Db | 1091 | AAGATCCACATTCCTCCATCATTCGATCCGTGTCCGAGCATCAGCCACACCTTGGTTCCTCG | 1150 | |
| Qy | 361 | PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTYrCysIle | 380 | |
| Db | 1151 | TTCITCTTGATCTGCACATCCTGGTGTGGCGCTTCCCACTGGGAGTGTGGTACTGCATC | 1210 | |
| Qy | 381 | LysGlnIleAsnaspGluArgValPheValValLeuTYrAlalleSerAlaValTYrPhe | 400 | |
| Db | 1211 | AAGCAGATCAACGACGAGCGCGTTTTCTGTGTGCTGTACGCCATCATGTCGGTTTACTTC | 1270 | |
| Qy | 401 | AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly | 420 | |
| Db | 1271 | GCTGCTGTGATGTGTGCTGTTGATGTTTGCACCTCACGCCGGTGTGTGCTGTCGTCGCGGA | 1330 | |
| Qy | 421 | VallaIaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLYsArgMetGly | 440 | |
| Db | 1331 | GTGGCCCTTTTCGGACCTGTTGGATGTGTTCTTGCAAGAGGATTCGTCTAAGCAATGGCC | 1390 | |
| Qy | 441 | ThrAlalleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLYsLysThr | 460 | |
| Db | 1391 | ACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCATTTAGAACAAGACG | 1450 | |
| Qy | 461 | LeuTYrAspLYsAlaGlyLYsLeuLYsHisArgThrLYsHisAspAlaGlnGlnAspThr | 480 | |
| Db | 1451 | CTGTACGACAAGGCTGGCAGCTGAGACATCGTACTAAGCATGATGCCACGAGGATACT | 1510 | |
| Qy | 481 | GlyValSerSerAsnLeuLYsSerIleValIleLeuAlaValLeuMetLeuLeuMetMet | 500 | |
| Db | 1511 | GGCGTCACTCCAACCTGAAGAGATVTGTTATVTTTGGCCGTCTCTAAATGCTGTGTATGATG | 1570 | |
| Qy | 501 | PheAlaValHisCysThrTrpValThrSerAsnAlaTYrSerSerProSerIleValLeu | 520 | |
| Db | 1571 | TTCGCTGTCCACTGCAGTGGGTGACCAGCAATGCTACTCCAGTCCCTCCATTGCTTG | 1630 | |
| Qy | 521 | AlaPheHisAsnSerGlnaspGlySerArgAsnIleLeuAspAspPheArgGluAlaTYr | 540 | |
| Db | 1631 | GCTTTCCAACACAGTCAAGATGATCCCGACATTTTAGACGATTTTCAGAGAGGCTTAC | 1690 | |
| Qy | 541 | TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerrtpPheAspTYrGly | 560 | |

| | | | | |
|-------------------|--|---------------|--|-----------------|
| DB | 1691 | TACTGGCTTTCG | CAGAACACTCGCGGATGATGCTCGGCTTATGCTTGGTGGGATTAACGA | 1750 |
| QY | 561 | TyrGlnIle | laGlyMetAlaAsnArgThrThrIleuValAspAsnAsnThrTrpAsnAsn | 580 |
| DB | 1751 | TACCAGATAGCGG | AATGCAACAGAACGACGCTAGTGGATATATATACGTGGAAACAAT | 1810 |
| QY | 581 | SerHisIle | alaLeuValGlyIysAlaMetSerSerThrGluGluIysSerTrpGluIle | 600 |
| DB | 1811 | AGTCACATAGCG | TGGTGGCAAGCAATGCTTCACCCGAGGAGAAAGTCCCTACGAAAT | 1870 |
| QY | 601 | MetThrSer | IleuAspValAspTrpValLeuValIlePheGlyValIleGlyTrpSer | 620 |
| DB | 1871 | ATGACATCTCTT | CACGTGACATACGCTTTTGTGTGATCTTTGCGGTGTGATCGGTATTCT | 1930 |
| QY | 621 | GlyAsnAsp | IleAsnIysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys | 640 |
| DB | 1931 | GGCGATGATATC | AAACAAAGTCTCTGTGGATGGTCCGATTTGCTGAGGAGAGCATCCCAAG | 1990 |
| QY | 641 | AspIleIys | GluSerAspTrpPheThrAspArgGlyGluPheArgValAspAlaGluGly | 660 |
| DB | 1991 | GACATTAAGGAA | AGCGAATTAACCCAGCCGCGTGAATTCAGGGTAGATGCGGAAGGT | 2050 |
| QY | 661 | AlaProAla | IleuLeuAsnCysLeuMetTrpIysLeuSerTrpTrpArgPheGlyGluLeu | 680 |
| DB | 2051 | GCTCCGGCCCTG | CTCACTGCCTTATGTACAAATTAGCTACTACAGATTCGGGGNATTG | 2110 |
| QY | 681 | LysLeuAsp | TrpArgGlyProSerGlyTrpAspArgThrArgAsnAlaValIleGlyAsn | 700 |
| DB | 2111 | AAGTTGGACTAC | AGAGTCCATCTCGATATGATCGCACAGTAAACGCCCTCATTTGGGAAT | 2170 |
| QY | 701 | LysAspPhe | AspLeuThrTrpLeuGluGluAlaTrpThrThrGluHisTrpLeuValArg | 720 |
| DB | 2171 | AGGACTTCGATC | TGACTACTCTCGAGAGGACCTACACACAGAACACATGGCTGTGTGCG | 2230 |
| QY | 721 | IleTrpArg | ValIysIysProHisGluPheAsnArgProSerIleuIysThrIysGluArg | 740 |
| DB | 2231 | ATCTATAGGGTG | AAGAGCGCATGATGATTCATATAGACCATCACTGAAGACCAAGGAGAGA | 2290 |
| QY | 741 | ThrIlePro | ProAlaAsnPheIleSerArgIysAsnSerIysArgArgIysGlyTrpIle | 760 |
| DB | 2291 | ACGATTCCTCC | AGAAACTTCATTCGAGAAAGAACTCTAAGCGTCGCAAGGGCTACATA | 2350 |
| QY | 761 | ArgAsnArg | ProValValIysGlyIysArgThrIleuIys | 774 |
| DB | 2351 | CGAAACCGACG | CGGTGTGTTTAAGGGAAACGACCTTGAAA | 2392 |
| RESULT 3 | | | | |
| LOCUS | CQ589353 | | 2699 bp | DNA |
| DEFINITION | Sequence 17111 from Patent WO0171042. | | linear | PAT 02-FEB-2004 |
| ACCESSION | CQ589353 | | | |
| VERSION | CQ589353.1 | GI:41648215 | | |
| KEYWORDS | | | | |
| SOURCE | Drosophila sp. | | | |
| ORGANISM | Drosophila sp. | | | |
| | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | | |
| | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | | |
| | Ephydroidea; Drosophilidae; Drosophila. | | | |
| REFERENCE | | | | |
| AUTHORS | Venter, J.C., Adams, M., Li, P.W. and Myers, E.W. | | | |
| TITLE | Detection kits, such as nucleic acid arrays, for detecting the | | | |
| | expression of 10,000 or more Drosophila genes and uses thereof | | | |
| JOURNAL | Patent: WO 01/042-A 17111 27-SEP-2001; | | | |
| | PE Corporation (NY) (US) | | | |
| FEATURES | | | | |
| source | Location/Qualifiers | | | |
| | 1..2699 | | | |
| | /organism="Drosophila sp." | | | |
| | /mol_type="unassigned DNA" | | | |
| | /db_xref="taxon:7242" | | | |
| ORIGIN | | | | |
| Alignment Scores: | 3..626-314 | | | 2699 |
| Pred. No.: | | | | |

Score: 4046.00 Matches: 774
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-028-384-8 (1-774) x CQ589353 (1-2699)

| | | | |
|----|-----|--|------|
| QY | 1 | MetAsnArgThrProGlyMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr | 20 |
| DB | 80 | ATGAATCGGACGCCGCAAGATGCTGAACAGCAAGGTGGCTGCTACAGAGCCCTAATCACC | 139 |
| QY | 21 | PheAlaIleLeuLeuIleAlaThrPheAlaGlyPheSerSerArgLeuPheAlaValle | 40 |
| DB | 140 | TTGCGCATCTCTGCTAATCGCTGGCTGGCCGGATTTCTCTCGCCCTCTTGGCCGCTATC | 199 |
| QY | 41 | ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrA:GAlaThrAlaTyr | 60 |
| DB | 200 | CGTTTCGAGTGGATATTCATGAGTTTGATCCGTGGTTTCACTACCGGGCCACCGCCTAC | 259 |
| QY | 61 | MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluAlaTyrPro | 80 |
| DB | 260 | ATGGTGCAAGATGGTGTGTAACAATCTCTCAACTGGTTCGACGAGCGCATGCTATCCG | 319 |
| QY | 81 | LeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle | 100 |
| DB | 320 | CTCGGAGATTTGGGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGGCGGAATC | 379 |
| QY | 101 | HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu | 120 |
| DB | 380 | CATTGGCTGTGTCAGTACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG | 439 |
| QY | 121 | AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp | 140 |
| DB | 440 | CGCGCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG | 499 |
| QY | 141 | SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer | 160 |
| DB | 500 | TCCCGGGCGCGGCTCTTCGCGCGAGCTTCATCGCCATCGTGGCTGCTGCTACATCAGT | 559 |
| QY | 161 | ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr | 180 |
| DB | 560 | AGTTCGGTGGTGGATCGTACGATPACGAGGGGCAATGCCATATTCGCCCTGCAAGTTCAC | 619 |
| QY | 181 | TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaA | 200 |
| DB | 620 | TACTTCCTGCTGGTGGCTCAGTGAACATGGATCGGTGTTCTGGTGGCGGAGCGGCT | 679 |
| QY | 201 | LeuSerTyrPheThrMetValSerAlaTrpGlyGlyTyrValPheIleLeuLeuIle | 220 |
| DB | 680 | TTGTCCTACTTCTACATGGTGGTGGCTGGGCTGGCTGCTGCTGCTGCTGCTGCTGCT | 739 |
| QY | 221 | ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer | 240 |
| DB | 740 | CCCTCGCACCTCTTCGTACTGCTCATATGGGCGAGGTACTCGCGGCTGCTGCTGCTGCT | 799 |
| QY | 241 | TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe | 260 |
| DB | 800 | TACAGCACCTTACATCTCTGGGACTGCTGCTTCTCCATGCGATCCCTTCGTGGGATTC | 859 |
| QY | 261 | GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla | 280 |
| DB | 860 | CAACCATACGACACAGTGAACATGGCTGGCTGGGAGTGTGTGTCTCTCTTATGGCC | 919 |
| QY | 281 | ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe | 300 |
| DB | 920 | GTGGCCACCTTGGCCATTTGAGTCCGTGCTGCTGCGCAACAGAGTTCGCGAGCTGTC | 979 |
| QY | 301 | IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet | 320 |
| DB | 980 | ATCGTGGCGGATTTGTGGTGGCGGTGGCGCTTTTGTGGCGGCTGCTGCTGCTGCTGCT | 1039 |
| QY | 321 | LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla | 340 |

| | | | |
|----|------|---|------|
| DB | 1040 | CTGGGGCTTGTGGCCCGGTGGAGTGAACGCTTCTACTCGTGGGATATGCTGCTACGCC | 1099 |
| QY | 341 | LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer | 360 |
| DB | 1100 | AAGATCCATATCCCATCATTCATCGTGTGGAGCATCAGCCACCATCTGGTTCTCG | 1159 |
| QY | 361 | PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle | 380 |
| DB | 1160 | TTCTTTCTTGTATCTGCATCTCTGGTGGCTTCCAGTGGAGTGGTACTGCATC | 1219 |
| QY | 381 | LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe | 400 |
| DB | 1220 | AAGCAGATCAACGACGAGCGCTTTTCGTGGTGTGTCAGCCATCAGTGGCTTTACTTC | 1279 |
| QY | 401 | AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly | 420 |
| DB | 1280 | GCTGGTGTGATGGTGGTGTGATGTGACCTTCAGCCGCGTGTGCTGCTGCTGCTGCTG | 1339 |
| QY | 421 | ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly | 440 |
| DB | 1340 | GTGGCCTTTTCGGGACTGTTGATGTGTTCTCTGCAAGAGGATTCCTTAAGCAGATGGC | 1399 |
| QY | 441 | ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr | 460 |
| DB | 1400 | ACAGCCATTAAGCGCAGCCACCGAAGTGAAGCTGAGGATTCATTCAGAAAGAACG | 1459 |
| QY | 461 | LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr | 480 |
| DB | 1460 | CTGTACGACAAAGCTGGCAGCTGAAGCATCGTACTAAGCATGATGCCAGCAGGATACT | 1519 |
| QY | 481 | GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet | 500 |
| DB | 1520 | GGCGTCAGCTCCAACTGAAGAGTATCTGTTATTTGGCGCTTCTAATGCTGTGTGATG | 1579 |
| QY | 501 | PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu | 520 |
| DB | 1580 | TTGCGTGTCCACTGACGCTGGGTGACCAAGTGCCTACTCCAGTCCCTCCATGCTGTG | 1639 |
| QY | 521 | AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr | 540 |
| DB | 1640 | GCTTTCCACAAACAGTCAAGATGGATCCCGCAACATTTTAGACGATTTTCAGAGAGCTTAC | 1699 |
| QY | 541 | TyrTrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTrpAspTyrGly | 560 |
| DB | 1700 | TACTGGCTTTCCGAGAACACTGCGCATGATGCTCGCGTATGCTGTGCTGGGATTCGGA | 1759 |
| QY | 561 | TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnThrTrpAsnAsn | 580 |
| DB | 1760 | TACCAGATAGCGGGAATGGCAACAGACGCTAGTGGATATATATACGTGGAACAT | 1819 |
| QY | 581 | SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle | 600 |
| DB | 1820 | AGTCACATAGCGCTGGTTGGCAAGCAATGCTTCAACCGGAGGAGAGTCTACGAAAT | 1879 |
| QY | 601 | MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer | 620 |
| DB | 1880 | ATGACATCTCTGACGTGGACTACGTTTGGTGAICTTTGGCGGTGATCGGCTATCT | 1939 |
| QY | 621 | GlyAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys | 640 |
| DB | 1940 | GGCGATGATATCAACAAGTTCCTGTGGATGCTCGAATTCGTAGGGAGAGATCCCAAG | 1999 |
| QY | 641 | AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly | 660 |
| DB | 2000 | GACATTAAGAAACGATTTACTTTTACCGACCGCGTGAATTCAGGGTAGATGCCAAGGT | 2059 |
| QY | 661 | AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu | 680 |
| DB | 2060 | GCTCGGCGCTGCTCAACTGCTTATGTACAAATTAAGCTACTACACATTCGGGGAATG | 2119 |
| QY | 681 | LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn | 700 |

Db 2120 AAGTGGACTACAGAGGTCCATCTGGATATGATCGCACACGTAAACGCGCTCATTTGGGAAT 2179
 QY
 QY 701 LysAspPheAspLeuThrTyrluGluGluAlaTyrluThrGluHisTrpLeuValArg 720
 Db 2180 AAGGACTTCGATCTACCTACCTGGAGAGGCTTACACACAGAACACTGGCTTGTTCGC 2239
 QY
 QY 721 IleTyArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 Db 2240 AICTATAGGTGAGAGCGCATGAGTTCAATAGACCATCACTGAAGACCAAGGAGAGA 2299
 QY
 QY 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrlle 760
 Db 2300 ACGATTCTCCAGCAACTTCTATTCGAGAAAGAACTCTAAGCGTCGCAAGGGCTACATA 2359
 QY
 QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 Db 2360 CGAAACCGACCGGTGTGTTAAGGGAAACGAACTTGAA 2401
 RESULT 4
 CQ589352/c
 LOCUS 4922 bp DNA linear PAT 02-FEB-2004
 DEFINITION Sequence 17110 from Patent WO0171042.
 ACCESSION CQ589352
 VERSION CQ589352.1 GI:41648214
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila sp.
 Drosophila sp.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 1 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
 Detection kits, such as nucleic acid arrays, for detecting the
 expression of 10,000 or more Drosophila genes and uses thereof
 Patent: WO 017042-A 17110 27-SEP-2001;
 PE Corporation (NY) (US)
 FEATURES
 source
 1..4922 Location/Qualifiers
 /organism="Drosophila sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:7242"
 ORIGIN
 Alignment Scores: 1..28e-303 Length: 4922
 Pred. No.: 3918.50 Matches: 772
 Score: 90.93% Conservative: 0
 Percent Similarity: 90.93% Mismatches: 2
 Best Local Similarity: 96.85% Indels: 77
 Query Match: 6 Gaps: 4
 DB: 6
 US-10-028-384-8 (1-774) x CQ589352 (1-4922)
 QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrlSerSerLeuIleThr 20
 Db 3843 ATGAATCGGACCGCGAAGATGCTGAACAGCAAGGTGGCTGGCTACAGACGCTTAATCAC 3784
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 Db 3783 TTCGCCATCCGCTAATCGCTGGCTGGCCGGATTTTCCTCTCGCTCTTCGCCGTCATC 3724
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAlaThrAlaTyrl 60
 Db 3723 CGTTTCGAGTCGATTAATCCATGAGTTTGATCCGCTGGTTCAACTACCGGGCCACCGCTAC 3664
 QY 61 MetValGlnAsnGlyTrpTyrlAsnPheLeuAsnTrpPheAspGluAlaTrpTyrlPro 80
 Db 3663 ATGGTGCAAGATGGTGTACACTTCTCACTGGTTCCAGAGCGCGCATGGTATCCG 3604
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrlProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 3603 CTCGGCAGGATTTGGGGCGGTACCGTCTATCCCGCGCTGATGATTACGTCGGCGGAATC 3544

QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 3543 CATTGGCTGCTGCACGTAATCAACATACCGGTCCATATTCGTGACATCTCGTGTTCCTG 3484
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrlLeuLeuThrLysGluLeuTrp 140
 Db 3483 GGGCGGATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGGTGG 3424
 QY 141 SerAlaGlyAlaGlyLeuPheAlaIleAspPheIleAlaIleValProGlyTyrlSer 160
 Db 3423 TCCGCGGGCGCGGCTCTTCGCCCGCAGCTTCATCGCATGCTGCTGGCTACATCAGT 3364
 QY 161 ArgSerValAlaGlySerTyrlAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 Db 3363 AGTCCGCTGGCTGGATCGATGATACAGAGGCGATTCGATATTCGCCCTCGAGTTCACC 3304
 QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAla 200
 Db 3303 TACTTCTGTGGTGGCTCAGTGAAGACTGATCGGTCTTCTGTGCGCGCAGCCGCT 3244
 QY 201 LeuSerTyrlPheTyrlMetValSerAlaTrpGlyTyrlValPheIleIleAsnLeuIle 220
 Db 3243 TTGTCTTACTTCTACATGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGATA 3184
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrlSerProArgLeuLeuThrSer 240
 Db 3183 CCCTCGACGCTCTTCGTACTGCTCATTTATGGCGAGGTACTCGCCGCTGCTGCTACCCAGC 3124
 QY 241 TyrlSerThrPheTyrlIleGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 Db 3123 TACAGCACCTTCTACATCTCGGACTGCTGTTCTCCATGCAGATCCCTTCCTGGGATTC 3064
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 Db 3063 CAACCGATAGCACAGTACACATGGCTGGCTGGAGTGTGTGCTGCTTATGGCC 3004
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 Db 3003 GTGGCCACCTTGGCGCATTTGCGAGTCCGCTGTGCGCGAAGAGTTCGGAGAGCTGTTC 2944
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet 320
 Db 2943 ATGTCGCGCGGATTTGCTGGTGGCGCTTGGCTGCTTGTGGCGCTGCTGCTGCTGCTGCTG 2884
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrlSerLeuTrpAspThrGlyTyrlAla 340
 Db 2883 CTGGCGCTTGTGGCGCTGGAGTGGAGCGCTTCTACTCGCTGGGATATCTGGCTACCGCC 2824
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 Db 2823 AAGATCCACATTCCTCATTCGCTGCGAGCATCAGCCACCCACCATCTGGTTCGTG 2764
 QY 361 PhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrlCysIle 380
 Db 2763 TTCTTCTTGTATCTGCACATCTTGTGTGGCGCTTCCAGTGGAGTGTGGTATCTGCTCATC 2704
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrlAlaIleSerAlaValTyrlPhe 400
 Db 2703 AAGCAGATCAACAGCAGCGCTTTCGTGGTGTGCTGACGCCATCAGTGGCGTTCCTTC 2644
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 2643 CTTGGTGTGATGGTGGCTTGTATGCTGACCTCACCGCGCTGCTGCTGCTGCTGCTGCTG 2584
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnLysAspSerSerLysArgMetGly 440
 Db 2583 GTGGCGCTTTCGGAGCTTGTGGATGCTTCTCTCAAGAGGATTCGCTAAGCGGATGGCC 2524
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 Db 2523 ACAGCCATACCGCAGCCACCGAGTGGATGAGTCAAGCTCAGGATTCATTTGAGAGAGACG 2464
 QY 461 LeuTyrlAspLysAla----- 465

```

1385 CAATCTGATCTTTGTCAGAACTCTTAAGCGTCGACGGGCTCATATACGAAACGACCGGTT 1326
766 ValVallysGlyLysArgThrLeuLys 774
1325 GTTGTAAAGGAAACGACCTTGAA 1299

RESULT 5
AC007853/c 162921 bp DNA linear INV 27-FEB-2001
Drosophila melanogaster, chromosome 3R, region 96B-96C, BAC clone
BACR03L02, complete sequence.
AC007853
VERSION AC007853
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 162921)
Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banzone, J., Beeson, K.V., Busam, D.A.,
Carlson, J.W., Center, A., Chape, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doupl, E., Doyle, C., Dresnek, D., Farfan, D.,
Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattai, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuncio, J.,
Paclob, J., Parag, V., Park, S., Patel, S., Pfeiffer, B.,
Phoenavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 3R, region 96B-96C
Unpublished
2 (bases 1 to 162921)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chape, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Humnasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 27, 2001 this sequence version replaced gi:5670481.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
Location/Qualifiers
1. 162921
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="96B-96C"
/clones="BACR03L02 (D766)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6)"

TITLE JOURNAL
REFERENCE AUTHORS
Drosophila melanogaster, chromosome 3R, region 96B-96C
Unpublished
2 (bases 1 to 162921)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chape, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Humnasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclob, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (17-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 27, 2001 this sequence version replaced gi:5670481.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.
Location/Qualifiers
1. 162921
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="96B-96C"
/clones="BACR03L02 (D766)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6)"

FEATURES
source
1. 162921
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="96B-96C"
/clones="BACR03L02 (D766)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6)"

ORIGIN
1385 CAATCTGATCTTTGTCAGAACTCTTAAGCGTCGACGGGCTCATATACGAAACGACCGGTT 1326
766 ValVallysGlyLysArgThrLeuLys 774
1325 GTTGTAAAGGAAACGACCTTGAA 1299

```

Alignment Scores:

Pred. No.: 1,16e-301 Length: 162921
 Scores: 3918,50 Matches: 772
 Percent Similarity: 90.93% Conservative: 0
 Best Local Similarity: 90.93% Mismatches: 2
 Query Match: 96.85% Indels: 77
 DB: 3 Gaps: 4

US-10-028-384-8 (1-774) x AC007853 (1-162921)

| | | | |
|----|--------|--|--------|
| QY | 1 | MetAsnArgThrProTyrMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr | 20 |
| DB | 108481 | ATGAATCGAGCGCGGAGATGCTGAACAGCAGAGTGCGCTACAGCAGCTAATACCC | 108422 |
| QY | 21 | PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValile | 40 |
| DB | 108421 | TTCCGCAATCGCTAATCGGCTGGCTGGCGGATTTTCCTCTCGCCCTCTCGCCGTCATC | 108362 |
| QY | 41 | ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr | 60 |
| DB | 108361 | CGTTTCAGTGCATATTCATAGAGTTGATCCGTGTTCACTTACCGGCGCACCCCTAC | 108302 |
| QY | 61 | MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro | 80 |
| DB | 108301 | ATGGTGCAAGATGGTTGGTCAACTTCTCACTGTTTCGACGAGCGGCAATGGTATCG | 108242 |
| QY | 81 | LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyIle | 100 |
| DB | 108241 | CTCGCGAGGATTTGGCGCGGTACCGTCTATCCCGGCTGATATTACGTCCGCGGGAATC | 108182 |
| QY | 101 | HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu | 120 |
| DB | 108181 | CATTGGCTGTGCACGTACTCAACATACCGGTCCAATTCGTGACATCTGCTGTCTCTG | 108122 |
| QY | 121 | AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIleGluLeuTrp | 140 |
| DB | 108121 | GGCGGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACTCTGACCAAGAGGTGG | 108062 |
| QY | 141 | SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer | 160 |
| DB | 108061 | TCCGCGCGCGCGGCTTTTCGCGCGCAGCTTCATCGCCATCGTGCCTGGCTACATCAGT | 108002 |
| QY | 161 | ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr | 180 |
| DB | 108001 | AGGTGGTGCTGATCGTACGATAACGAGGCGCATTTCCCATATTCGCTTCGCGTTCACC | 107942 |
| QY | 181 | TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaAla | 200 |
| DB | 107941 | TACTTCTGTGGGTGGCTCAGTGAAGACTGGATCGTGTCTGTGCGCGCGAGCGCT | 107882 |
| QY | 201 | LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleLeuAsnLeuIle | 220 |
| DB | 107881 | TTTGCTTACTTCACATGAGTGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT | 107822 |
| QY | 221 | ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer | 240 |
| DB | 107821 | CCCCGTGACGCTTTCGTACTGCTCATTATGGGCGAGTACTCGCGCTGCTGACACAGC | 107762 |
| QY | 241 | TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe | 260 |
| DB | 107761 | TACAGCACCTTCTACATCTGGAGCTGTGTTCTCCATCGATCCCTTGTGGGATTC | 107702 |
| QY | 261 | GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla | 280 |
| DB | 107701 | CAACCGATACGCACAGTGAACATCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT | 107642 |
| QY | 281 | ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgIleLeuPhe | 300 |
| DB | 107641 | GTGGCCACCTTGGCCATTTGCGAGTCCGTGCTGTCGCGCAACAGATTCGGAGCTGTT | 107582 |
| QY | 301 | IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet | 320 |
| DB | 107581 | ATCGTCGCGGATTTGCTGGTGGCGGTGGCGTCTTTGTGGCGCTGCTGTGCTCACCATG | 107522 |
| QY | 321 | LeuGlyValValAlaProTyrSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla | 340 |
| DB | 107521 | CTGGCGCTTGTGGCGCGCGGAGTGACGCTTCTACTCGCTGTGGGATCTGGCTACGCC | 107462 |
| QY | 341 | LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrTrpPheSer | 360 |
| DB | 107461 | AGATCCACATCCCATCATTCATCGCTGTGGAGCATCAGCCACCATCTGGTCTCG | 107402 |
| QY | 361 | PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle | 380 |
| DB | 107401 | TTCTTCTTTGATCTGCACATCTCTGTGTGGCGCTTCCAGTGGAGTGTGTACTGCATC | 107342 |
| QY | 381 | LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe | 400 |
| DB | 107341 | AAGCAGATCAACGACGAGCGGCTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 107282 |
| QY | 401 | AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly | 420 |
| DB | 107281 | GCTGGGT | 107222 |
| QY | 421 | ValAlaPheSerGlyLeuLeuAspValPheLeuGlnAspSerSerLysArgMetGly | 440 |
| DB | 107221 | GTGGCTTTTCGGGACTGT | 107162 |
| QY | 441 | ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysIleThr | 460 |
| DB | 107161 | ACAGCCATTAAGCGGACGACCCGAGTGGATGAAGTGGATTCATTTGAGAGAGACG | 107102 |
| QY | 461 | LeuTyrAspLysAla | 465 |
| DB | 107101 | CTGTACGACAGGT--GAGTTCTTACTAACACATCCATCCGATTTGATTTTAAATACA | 107043 |
| QY | 466 | -----GlyValLeuLeuHisArgThrLysHisAspAlaGlnGlnAspThr | 480 |
| DB | 107042 | TGCAATTTGCGAGCTGGGAGCTGAAGCATCTGTAAGCATGATGATGCCACAGGATACT | 106983 |
| QY | 481 | GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet | 500 |
| DB | 106982 | GGCGTCAGCTCCAACTGAAGATATTGTTATTTTGGCCGTTCTTAATGCTGTTGATG | 106923 |
| QY | 501 | PheAlaValHisCysThrTyrValThrSerAsnAlaTyrSerSerProSerIleValLeu | 520 |
| DB | 106922 | TTGCTGTCTCATCTGACGCTGGGTGACCAAGTCCCTACTCCAGTCCCTCCATGCTTG | 106863 |
| QY | 521 | AlaPheHisAsnSerGlnAspGly | 528 |
| DB | 106862 | GCTTTCCACACAGTCAAGATGGGTGAAGCACCAGTAGCATCATGCTCTTTTAAACCTA | 106803 |
| QY | 529 | -----SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr | 541 |
| DB | 106802 | TCTAACCTTTTCTCACATAGATCCCGCAACATTTTAGACGATTTACAGAGGCTTACTAC | 106743 |
| QY | 542 | TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGlyTyr | 561 |
| DB | 106742 | TGCGTTTCGCAACACATGCGGATGATGCTCGCTTATGCTTGTGGGATTTACGGATAC | 106683 |
| QY | 562 | GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer | 581 |
| DB | 106682 | CAGATAGCGGATGCGCAACAGACGCGTAGTGGATTAATATACGTGGACATAGT | 106623 |
| QY | 582 | HisIleAlaLeuValGlyLysAlaMetSerSerThrGluLysSerTyrGluIleMet | 601 |
| DB | 106622 | CACATAGCGCTGGTTGGCAAGCAATGTCTTCAACCGGAGGAGAGTCTTACGAAATATG | 106563 |
| QY | 602 | ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly | 621 |
| DB | 106562 | ACATCTCTTGACGTGCATCATGTTTGTGTGATCTTTGGCGGTGTGATCGCTATTCTGGC | 106503 |
| QY | 622 | AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp | 641 |
| DB | 106502 | GATGATATCAACAGTTCCTGTGGATGTCGCAATTCGTGAGGAGAGCATCCCAAGGAC | 106443 |

QY 642 lleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyVala 661
 Db 106442 ATTAAGGAAGCGATTACTTTACCGACCGGTGAATTCAGGGTAGATCGCGAAGGTGCT 106383
 QY 662 ProAlaLeuAenCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 Db 106382 CCGGCCCTTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATTTCGGGGAATTGAAG 106323
 QY 682 LeuAspTyrArg----- 685
 Db 106322 TTGGACTACAG-GTAAGCGGAACATTTCTCTCAGGTAGCGATGCACACTAACTGCCTGT 106264
 QY 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 Db 106263 TTCAGAGTCCATCTGGATATCATCGCACAGTAAACCGCTCATTTGGGAATTAAGGACTTC 106204
 QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
 Db 106203 GATCTGACCTACTCTGGAGGCGCTACACACAGAACCTGGCTTGTTCGCATCTATAGG 106144
 QY 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
 Db 106143 GTGAAGAAGCGCGATGATTCATATAGACCATCACTGAAGACCAAGGAGAGACGATTCCT 106084
 QY 744 ProAlaAsnPheIleSerArgLys----- 751
 Db 106083 CCAGCAACTCTCATTTGAGAAAGGTAGTACCTCGACCTCCCAATGAACCTTCATTAA 106024
 QY 752 -----AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
 Db 106023 CAATCTGATCTTTGCAAGACTCTAAGCGTGGCAAGGCTACATACAGAACCGCGTT 105964
 QY 766 ValValLysGlyLysArgThrLeuLys 774
 Db 105963 GTTGTTAAGGAAACGACCTTGAA 105937

RESULT 6

AC008206/c

LOCUS

AC008206 181132 bp DNA linear INV 24-FEB-2001
 Drosophila melanogaster, chromosome 3R, region 96B-96B, BAC clone
 BACR03115, complete sequence.

ACCESSION

AC008206

VERSION

AC008206.10

KEYWORDS

HTG

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 181132)

Celnikier,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,

Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,

Rogers,Y., An,H., Baldwin,D., Bonjorno,J., Beeson,K.Y., Busam,D.A.,

Carlson,J.W., Center,H., Chapple,M., Davenport,L.B., Dietz,S.M.,

Dodson,K., Dotsis,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,

Fertler,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,

Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,

Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,

McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,

Paclob,J., Paragay,V., Park,S., Patel,S., Pfeiffer,B.,

Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,

Stapleton,M., Strong,R., Swirskas,R., Tector,C., Williams,S.M.,

Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Sequencing of Drosophila chromosome 3R, region 96B-96B

Unpublished

2 (bases 1 to 181132)

Celnikier,S.E., Abayani,A., Arcania,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.

Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclob,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

TITLE

JOURNAL

COMMENT

Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Direct Submission
 Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Feb 24, 2001 this sequence version replaced gi:7208834.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.

FEATURES

source

1..181132
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="Y, cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="96B-96B"
 /clone="BACR03115 (D765)"
 /disease lib="RPC1-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACE3.6)"
 Alignment Scores:
 Pred. No.: 1.33e-301 Length: 181132
 Score: 3918.50 Matches: 772
 Percent Similarity: 90.93% Conservative: 0
 Best Local Similarity: 90.93% Mismatches: 2
 Query Match: 96.85% Indels: 77
 DB: 3 Gaps: 4
 US-10-028-384-8 (1-774) x AC008206 (1-181132)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
 Db 171813 ATGATCGACCGCGAAGATCTGAGACGACGAGTGGCTACACGACCTAATCACC 171754
 QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 Db 171753 TTCCGCTCTCTGCTAATCGCTGCTGCGCGGATTTTCTCTCGCTCTTCGCGCTCATC 171694
 QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
 Db 171693 CGTTTCGAGTCGATTATCCATGATGTTGATCCGTTCAACTACCGGCGCACGCGCTAC 171634
 QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrPro 80
 Db 171633 ATGTCGACGATGTTGGTACACTTCTCACTGTTCCAGCGCGCATGATGATTCG 171574
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 Db 171573 CTCGCGCAGGATTGTCGCGGTACCGTCTATCCCGCGCTGATGATTCGTCGCGGGAATC 171514
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 Db 171513 CATTCGCTGCTGCACGACTCAACATACCGGTCCATATTCGTGACATCTCGGTGTTCTG 171454
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 Db 171453 GCGCGGATCTTCAGTCGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGTGTGG 171394
 QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
 Db 171393 TCCGCGGCGCGCGCTCTTCGCGCGCAGCTTCATCGCCATCGTGCCTGCATCATCAGT 171334
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180

171333 AGGTCGGTGGTGGATCGTATACGAGGAGATTCGATATTCGCTCGAGTTCACC 171274
 181 TyrPheLeuThrValArgSerValLysThrGlySerValPheThrSerAlaAlaAla 200
 171273 TACTTCCTGGTGGTGGCTCAGTGAAGACTGGATCGGTCTGGTGGCGCGAGCGCT 171214
 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheLeuLeuLeu 220
 171213 TTGTCCTACTTCTACATGGTGTGGCGCTGGGCTGGGCTGGTGTTCATCAACCTGATA 171154
 221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
 171153 CCCCTGCACGCTTCTGACTGCTCATTTATGGGAGAGTACTCGCGGCTGCTGTCACGAC 171094
 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 171093 TACAGCACCTTCTACATCCTGGGACTGCTTCTCCATGCAGATCCCTTCGTGGGATTC 171034
 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 171033 CAACCGATACGACACGATGACACATGGCTGCGCTGGGAGTGTGTGCTCTTATGGCC 170974
 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 170973 GTGGCCACCTTGGCCCATTTGCAAGTCCGCTGCTCGCGCAACGAGTTCGGAAGCTGTC 170914
 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 170913 ATCGTCGGCGATGCTGCTGGGCGTGGGCTGCTTGTGGCGCTGCTGCTGCTCACCATG 170854
 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 170853 CTGGCGCTGTGGCGCGCTGGAGTGGAGCTTCTACTCGCTGTGGGACTCGCTGCTACGCC 170794
 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 170793 AAGATCCACATCCCATTCATTCATTCGCTGCGAGCATTCAGCCACCATTCGTTGCTTCG 170734
 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 170733 TTCTTCTTGTATCTGCATCTCTGGTGGCTGCGCTTCCAGTGGGAGTGTGTACTGCATC 170674
 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 170673 AAGCAGATCAACGACGAGCGGCTTTCGTGCTGCTGATCGCCATCAGTCCGCTTACTTC 170614
 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 170613 GCTGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170554
 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 170553 GTGGCCCTTTCGGACGCTTGGAGTGTCTCTCAAGAGATTCGCTTAAGCAATGGGC 170494
 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaAspSerIleGluLysLysThr 460
 170493 ACAGCCATAAGCGCAGCCAGCGAGTGAAGCTGAGCTGAGCTTCCATTCAGAGAAGACG 170434
 461 LeuTyrAspLysAla 465
 170433 CTGTACGACAGGT-GAGTTCCTTACTTAACACATCCATGGGATTTGATTTTAAATACA 170375
 466 -----GlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
 170374 TCGCATTTGAGGCTGGCAAGCTGAAGCATCGTACTTAAGCATGATGCCAGCAGGATACT 170315
 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMet 500
 170314 GGCTGAGCTTCCAACTTCAAGATATGTATTTTGGCGCTTCTAATGCTGTGTATGATG 170255
 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerProSerIleValLeu 520

Db 170254 TTCGCTGTCCACTGTCACGCTGGGTGACCAAGCAATGCTTACTCCAGTCCCTCCATTTGCTTG 170195
 Qy 521 AlaPheHisAsnSerGlnAspGly 528
 Db 170194 GCTTTCCACACAGTCAGATGGGTAAAGCACCAGTAGCATCATGTCTTCTTTAAACCTA 170135
 Qy 529 -----SerArgAsnIleLeuAspPheArgGluAlaTyrTyr 541
 Db 170134 TCTAACCTTTTCTCCACTAGATCCCGCAACATTTTAGACGATTTTCAGAGGCTTACTAC 170075
 Qy 542 TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGlyTyr 561
 Db 170074 TGGCTTTCCAGACACATCGCGATGATCTCGCTTATGCTTCTGCTGGATACGGATAC 170015
 Qy 562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 581
 Db 170014 CAGATAGCGGGAATGGCAACAGACGCTAGTGGATATAATAATACGTGGAAACATAGT 169955
 Qy 582 HisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMet 601
 Db 169954 CACATAGCGCTGCTGGCAAGGCAATGCTTCAACCGAGAGAGAGTCTACGAAATATG 169895
 Qy 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
 Db 169894 ACATCTCTTTGACGTGGACTACGTTTTTGGTGTATCTTTGGCGGTGTGATCTTCTGGC 169835
 Qy 622 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 641
 Db 169834 GATGATATCAACAAGTTCTCTGGATGTCGCAATGCTGAGGAGAGCATCCCAAGAC 169775
 Qy 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
 Db 169774 ATTAAGGAAGCGATTTACTTACCGACCGCGTGAATTCAGGGTAGATGCGAAGGTGCT 169715
 Qy 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 Db 169714 CCGGCGCTGCTCACTGCTTATGTACAAATTAAGCTACTACAGATTCGGGGATTTGAG 169655
 Qy 682 LeuAspTyrArg 685
 Db 169654 TTGGACTACAG-GTAAGCGGAAACATTTCTCTCAGTAGGATGCACACTAACTGCTGT 169596
 Qy 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 Db 169595 TTCAGAGGTCCATCTGATATGATCGACAGTAAAGCCGCTCATTTGGGATTAAGACTTC 169536
 Qy 704 AspLeuThrTyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArg 723
 Db 169535 GATCTGACCTACTCGAGAGGCGCTTACACACAGACACTGGCTTGTTCGATCTATAGG 169476
 Qy 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluAspThrIlePro 743
 Db 169475 GTGAAGAGCCCATGATTTCAATAGACCATCTGAGACCAAGAGAGAACGATTCCT 169416
 Qy 744 ProAlaAsnPheIleSerArgLys 751
 Db 169415 CCAGCAAACTTCATTTCCAGAAAGGTAGTGTACCCCTCGAGCTCCCAATGAATTCATTAA 169355
 Qy 752 -----AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
 Db 169355 CAATCTGATCTTCTCAAGATCTTAAGCTGCAAGGCTACATACGAAACCGAGCGGT 169296
 Qy 766 ValValLysGlyLysArgThrLeuLys 774
 Db 169295 GTTGTAAAGGAAACGAACTTGA 169269

RESULT 7
 AE003750/c
 LOCUS
 DEFINITION
 Drosophila melanogaster chromosome 3R, section 88 of 118 of the complete sequence.
 ACCESSION
 AE003750 AE002708 AE014297 AE002644
 VERSION
 AE003750.2 GI:23172229

227219 bp DNA linear INV 15-MAR-2004

KEYWORDS
SOURCE
ORGANISM

Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
Sutton, G.G., Wortman, J.R., Yeung, M., Zeng, Q., Chen, L.X.,
Brannon, R.C., Rogers, Y.H., Blazer, R.G., Chapple, M., Pfeiffer, B.D.,
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,
Abril, J.F., Agbayani, A., An, H.J., Andrews-Frank, C., Baldwin, D.,
Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,
Beckstead, R., Benson, P., Berman, B.P., Bhandari, D., Bolshakov, S.,
Borkov, D., Botchan, M.R., Bouck, J., Brockstein, P., Brothier, P.,
Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,
Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,
Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I.,
Dietz, S.M., Dodson, K., Dunn, P., Durbin, K.J., Evans, M., Dugan-Rocha, S.,
Dunkov, B.C., Dunn, P., Durbin, K.J., Evans, M., Dugan-Rocha, S.,
Fierera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodek, A., Gong, P., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennis, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Krivitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Murray, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nuskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Stapleton, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Simpson, M., Strong, R., Sun, E., Svirkas, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A.,
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodard, G.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, P.N., Zhong, M., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 227219)
Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W.,
Halpern, A., Patel, S., Adams, M., Champagne, M., Dugan, S.P., Frise, E.,
Hodgson, A., George, R.A., Hoskins, R.A., Laverly, T., Muzny, D.M.,
Nelson, C.R., Pacle, J.M., Park, S., Pfeiffer, B.D., Richards, S.,
Sodergren, E.J., Svirkas, R., Tabor, P.E., Wan, K.H., Stapleton, M.,
Sutton, G.G., Venter, C., Weinstock, G., Scherer, S.E., Myers, E.W.,
Gibbs, R.A. and Rubin, G.M.
Finishing a whole-genome shotgun: release 3 of the *Drosophila*
melanogaster euchromatic genome sequence
Genome Biol. 3 (12), RESEARCH0079 (2002)
22426065
12537568

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

3 (bases 1 to 227219)
Mitra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S.,
Hradecky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochuk, S.E.,
Smith, C.D., Tupy, J.L., Whitfield, E.J., Bayraktaroglu, L.,
Berman, B.P., Bettencourt, B.R., Celniker, S.E., de Grey, A.D.,
Dysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J.,
Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M.,
Rubin, G.M. and Lewis, S.E.
Annotation of the *Drosophila melanogaster* euchromatic genome: a
systematic review
Genome Biol. 3 (12), RESEARCH0083 (2002)
22426069

TITLE
JOURNAL
MEDLINE

PUBMED
REFERENCE
AUTHORS

12537572
4 (bases 1 to 227219)
Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J.,
Svirkas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E.,
Rubin, G.M., Ashburner, M. and Celniker, S.E.
The transposable elements of the *Drosophila melanogaster*
euchromatin: a genomics perspective
Genome Biol. 3 (12), RESEARCH0084 (2002)
22426070
12537573
5 (bases 1 to 227219)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
6 (bases 1 to 227219)
FlyBase
Direct Submission
Submitted (06-SEP-2002) University of California Berkeley, 539 Life
Sciences Addition, Berkeley, CA 94720, USA
7 (bases 1 to 227219)

REFERENCE
AUTHORS
CONSTRM
TITLE
JOURNAL
COMMENT
FEATURES
source

1. 227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement (149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96B10-96B11"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RB"
/db_xref="FLYBASE:FBgn0029157"
complement (join(875..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..5920))
/gene="ssh"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PA"
/protein_id="AAF56372.3"
/db_xref="GI:23172230"
/db_xref="FLYBASE:FBgn0029157"

gene

1. 227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement (149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96B10-96B11"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RB"
/db_xref="FLYBASE:FBgn0029157"
complement (join(875..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..5920))
/gene="ssh"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PA"
/protein_id="AAF56372.3"
/db_xref="GI:23172230"
/db_xref="FLYBASE:FBgn0029157"

mRNA

1. 227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement (149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96B10-96B11"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RB"
/db_xref="FLYBASE:FBgn0029157"
complement (join(875..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..5920))
/gene="ssh"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PA"
/protein_id="AAF56372.3"
/db_xref="GI:23172230"
/db_xref="FLYBASE:FBgn0029157"

mRNA

1. 227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement (149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96B10-96B11"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RB"
/db_xref="FLYBASE:FBgn0029157"
complement (join(875..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..5920))
/gene="ssh"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PA"
/protein_id="AAF56372.3"
/db_xref="GI:23172230"
/db_xref="FLYBASE:FBgn0029157"

CDS

1. 227219
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
complement (149..6378)
/gene="ssh"
/locus_tag="CG6238"
/map="96B10-96B11"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..6378))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RA"
/db_xref="FLYBASE:FBgn0029157"
complement (join(149..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5036..5341))
/gene="ssh"
/locus_tag="CG6238"
/product="CG6238-RB"
/db_xref="FLYBASE:FBgn0029157"
complement (join(875..3227,3313..3582,3649..3843,
3907..3972,4265..4656,4722..4967,5864..5920))
/gene="ssh"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PA"
/protein_id="AAF56372.3"
/db_xref="GI:23172230"
/db_xref="FLYBASE:FBgn0029157"

```

POQVMTSLKLTOTSHLGKSVSGSSGNIDRRSDSCSDVFSQVDSVFAKDBGEX
RQRKTHSNWTESLPGSGVLDPTPOQKQKQSNAILRPRGTQRRLPRRHASWSGSDN
KCLQKLEESTSLKRCQKQESQELLLEADAGRCPCFLSYRSASASAPRQRQPLRNC
SEELHSTGTENKNTTPGDEHATVILRVQQTLDADQRTSGNVQLKONFRAKHVVG
TGGGGGTGAAGSGSSSTIATSVAVAGASANPGKTTSHQSLPSSPVAGRAHVSA
ASNSNSASNSDSSSVSNCNDRNVKLVHKYQTTSCSSQAAAATPTVACAPLPAGTA
SNISHSCTTTTTSFKRPLSYDRSLRSHSTHSDSRPPAPRVVSHSSNVQHQQQ
QQOQLOEQHQOQVQIRRLAQHCKTHPLSLNAPKTFGNTAAAYNTM"
Complement (join (875..3227,3313..3582),3649..3843,
3907..3972,4265..4656,4722..4967,5036..5095))
/gene="ssr"
/locus_tag="CG6238"
/codon_start=1
/product="CG6238-PB"
/protein_id="AA014027.1"
/db_xref="GI:23172231"
/db_xref="FLYBASE:FBgn0029157"
/translation="MSQVCSLFDLILFLSIFPRDSEBDDSGNSKGNDRSECEFPAGK
GTALVALKDIPLPTOSERLSDTSRNSSTQNSNDIQLHLQSMFYLLQREPTLKM
AVKLESQRNRTLYLVIASRCCSGTSRRRHRIMRHSVRVGSAGTKSTSPAPV
TQRLSVEQATASAKCDKADENATAGDNKNTSGMEESCLIGDNCERTITGLV
VPIADTTIHLDDGGGFSVKVYEKTHIFKPSVQAMWALOTLHKVSKARENFIYAS
GPSHDLSSYERRIESDSCLENNAMDALERRPPSPDAIRNKPKKEETESVTKMK
LKAIMSVLDDEVTSKYRCGLFELIDMDLGEVKSFI DAEMLVILGQMDAPTKFEHV
YLGSWNASNLELQKNGVRLHNLNTRDINFPFGIFEYFNVRVYDDEKTLNKKWD
TFRITRKAAGSKVIVHCVGVSRSASVVIAYAMKAYQWFEQALEHVKRRCCKIP
KNFLNQLGYTSGMLDAMKXELQSKSETNLKSTDKDALPGESEPTPLIQAQNOAK
SKSTGEAGVTPDGEEDGSMRHSIAQSRMRVRSSTSPKTQAVTKQOOSQM
ENLTPERSVAEPKNNRFGSGNGENSVTCNOVLHIQKHTPLSVTRTHDLEAHRADQ
LPQOPVMTSLKLTOTSHLGKSVSGSSGNIDRRSDSCSDVFSQVDSVFAKDBGE
KROKRTHTSSGSGVLDPTPOQKQKQSNAILRPRGTQRRLPRRHASWSGSDN
NRCLPORTSSGVSYSNNRTALFEGVLODLKSSNCNVIEGVAVAVEPIVUGEGT
VKEFKOLESTSLKRCQKQESQELLLEADAGRCPCFLSYRSASASAPRQRQPLRC
NSEELMTSDIENKNTTPGDEHATVILRVQQTLDADQRTSGNVQLKONFRAKHV
GTGGGGGTGAAGSGSSSTIATSVAVAGASANPGKTTSHQSLPSSPVAGRAHV
ASNSNSASNSDSSSVSNCNDRNVKLVHKYQTTSCSSQAAAATPTVACAPLPAGT
ASNTSHSCSTTTTSFKRPLSYDRSLRSHSTHSDSRPPAPRVVSHSSNVQHQQQ
QQOQLOEQHQOQVQIRRLAQHCKTHPLSLNAPKTFGNTAAAYNTM"
6683..8846
/locus_tag="CG13645"
/map="56B11-96B11"
/db_xref="FLYBASE:FBgn0039254"
/join(6683..6899,6972..7378,7430..7531,7590..7741,
8330..8600,8709..8846)
/locus_tag="CG13645"
/product="CG13645-RA"
/db_xref="FLYBASE:FBgn0039254"
/join(6683..6899,6972..7378,7430..7531,7590..7741,
US-10-028-384-8 (1-774) x AE003750 (1-227219)

Alignment Scores:
Pred. No.: 1,78e-301 Length: 227219
Score: 3918.50 Matches: 772
Percent Similarity: 90.93% Conservative: 0
Best Local Similarity: 90.93% Mismatches: 2
Query Match: 96.85% Indels: 77
Dbs: 3 Gaps: 4

US-10-028-384-8 (1-774) x AE003750 (1-227219)

Qy 1' MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerLeuLeuThr 20
Db 131000 ATGAATCGAGCGCGCAAGATGCTGAACAGCAAGCGTGGCTGCTACAGCAAGCTAATCACC 130941
Qy 21 PheAlaLeuLeuLeuLeuAlaTyrPheLeuAlaGlyPheSerArgLeuPheAlaValle 40
Db 130940 TTCCGCATCTGCTAATCGCTGCTGCGCGGATTTCTCTCGCTCTCTTCGCGCTGCTAC 130881
Qy 41 ArgPheGluSerLlelleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr 60
Db 130880 CGTTTCGAGTCGATATCCATGAGTTTGATCCGCTGCTTCAACTACCGGGCCACCGCTAC 130821
Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrTyrPro 80

```

```

130820 ATGGTSCAGAAATGGTGGTACAACTTCTCAACTGGTTTCGAGAGCGGCGATGATCCG 130761
Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyLe 100
130760 CTGGCAGGAGTGTGGCGGCTACCGTCTATCCGCGCTGATGATTCGTCGGGCGGAATC 130701
Qy 101 HisTyrPheLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
130700 CATTGGCTCTGCACGTACTCAACATACGGTCCATATTCGTGACATCTCGGTGTTCTCG 130641
Qy 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr 140
130640 GCGCCGATCTTCAGTGGCGCTGACCTCCATCTCCACTACCTGCTGACCAAGAGCTGTGG 130581
Qy 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
130580 TCCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 130521
Qy 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
130520 AGTCTCGTGGCTGGATCGTACGATACGAGGCGCATTCGCATATTCGCGCTGACGTTCA 130461
Qy 181 TyrPheLeuTyrValArgSerValLysThrGlySerValPheTyrPheSerAlaAlaAla 200
130460 TACTTCTCTGGTGGCTGCGCTCAGTGAAGCTGGATCCGTGTTCTGGTGGCGCGAGCGCT 130401
Qy 201 LeuSerTyrPheTyrMetValSerAlaTyrGlyGlyTyrValPheIleIleAsnLeuLe 220
130400 TTGTCCTACTTCTACATGGTGTCCGCTCGGCTGGCTGCTACGTTGTTTCATCATCACT 130341
Qy 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
130340 CCCTTCGACGTCTTCGTACTGCTCATTTATGGCAGTACTCGCGCGTCTGCTGACCCAG 130281
Qy 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
130280 TACAGCACCTTCTACATCTCGGACGCTGCTGTTCTCCATCGAGATCCCTCTCGTGGGAT 130221
Qy 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
130220 CAACCGATACGACACGATGACATGCTGGCTGGGAGTGTGTTGCTCTTATGCGC 130161
Qy 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
130160 GTGGCCACCTTCGCGCATTTTGCAGTCCGTGCTGTCGCGCAACGAGTTCGCGAAGCTTC 130101
Qy 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet 320
130100 ATGCTCGCGGATGCTGGTGGCGCTTGGCGCTTTGTCGCGCTGCTGGTGTGCTCACC 130041
Qy 321 LeuGlyValAlaProTyrPheSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAla 340
130040 CTGGCGGTGTGGCGCGCTGGAGTGGAGCTTCTACTCGCTGTCGGATCTGGGTACGCC 129981
Qy 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrPheSer 360
129980 AAGATCCATATCCCATTCATTCGCTGTCGAGCATCAGCCACCATCTGTTCTCG 129921
Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrTyrCysIle 380
129920 TTCTTTCTGATCTGCATCTCTGCTGTCGCTTCCCGAGTGGAGTGTGGTACTGCTATC 129861
Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyr-Phe 400
129860 AAGCAGATCAACGACGAGCGCTTTTCGTGGTGTGTGTCGTCAGCCATCAGTGGGTTTACT 129801
Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
129800 GCTGGTGTGATGCTGCTTGTGATGTCCTCAGCGCGGTGGTGTGATGCTGCGCGGA 129741
Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440

```

```

Db 129740 GTGGCCTTTTCGGGACTGTGTGGATGTCTCTGCAAGAGGATTCGTCTAAAGCGAATGGC 129681
Qy 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
Db 129680 ACAGCCATAAGCGCAGCCACCGAAGTGGATGAAGCTGAGGATTCCTATTGAGAGAGACG 129621
Qy 461 LeuTyrAspLysAla 465
Db 129620 CTGTACGCAAGGT-GAGTCTTCTACTAAACACATCCATCCATGGATTTGATTTTAAATPACA 129562
Qy 466 -----GlyLysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThr 480
Db 129561 TCGCATTTTCAGGCTGGCAGCTGAAGCATCGTACTAGCATGATGCCAGCAGGATACT 129502
Qy 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
Db 129501 GCGGTGAGTCCCAACTGAAGAGTATGTATTTTGGCCGCTTCAATGCTGTGTGATGATG 129442
Qy 501 PheAlaValHisCysThrTyrValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
Db 129441 TTGCGTGTCCACTGCACGTGGGTGACACGAGTCCCTACTCCAGTCCCTCCATTTGCTTG 129382
Qy 521 AlaPheHisAsnSerGlnAspGly 528
Db 129381 GCTTTCCAAACAGTCAAGATGGTGAAGCACCAGTAGCATGTCTCTTTAAACCTA 129322
Qy 529 -----SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr 541
Db 129321 TCTAACCTTTTCTCCACTAGATCCCGCAACATTTTAGACGATTTTCAGAGGCTTACTAC 129262
Qy 542 TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGlyTyr 561
Db 129261 TGGCTTTTCAGACACACTGCCGATGATGTCGCTTATGTCTTGGTGGGATTTACGGATAC 129202
Qy 562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 581
Db 129201 CAGATAGCGGGATGGCAACAGACAGCGTAGTGGATATATATATATGTCGACAAATAGT 129142
Qy 582 HisIleAlaLeuValGlyLysAlaMetSerThrGluGluLysSerTyrGluIleMet 601
Db 129141 CACATAGCGCTGTTGGCAAGCAATGTCTTCAACCGAGGAGAGATCTCTACGAAATATG 129082
Qy 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
Db 129081 ACATCTTTGAGTGGACATGATTTTGGTATCTTTGGCGGTGTGATCGGCTATCTGGC 129022
Qy 622 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 641
Db 129021 GATGATATCAACAAGTCTCTGTGGATGTGTCGAAATTTGCTGAGGAGAGCATCCCAAGGAC 128962
Qy 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
Db 128961 ATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGTTAGATGCCGAAGGTCT 128902
Qy 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
Db 128901 CCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAG 128842
Qy 682 LeuAspTyrArg 685
Db 128841 TTGGACTACAG-GTAAAGCGGAACATTTCTCTCAGGTAGCGATGACACTAACTGCTGCTGT 128783
Qy 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
Db 128782 TTCAGAGGTCCTCATCTGATATGATCGCACGATGACCGCGTCATTTGGGAATTAAGGACTTC 128723
Qy 704 AspLeuThrTyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArg 723
Db 128722 GATCTGACCTTACTCGAGAGGCCCTACACACAGAACACTGCTGCTTGTTCGATCTATAGG 128663
Qy 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluAlaThrIlePro 743
Db 128662 GTGAAGAGCGCGATGAGTTCAATGACCATCACTGAGACCAAGGAGAGAACGATTCCT 128603

```

744 ProAlaAsnPheIleSerArgLys-----751

Db 128602 CCAGCAAACTTCAITTCGAGAAAGGTAGTGTACCTGCAGCTCCCAATGAACCTTCAATA 128543

Qy 752 -----AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765

Db 128542 CAATCTGATTTTTCGAGAACTTAAGCGTCGCAAGGGCTACATACGAAACCCGCGTT 128483

Qy 766 ValValLysGlyLysArgThrLeuLys 774

Db 128482 GTTGTGAAGGAAACCAACCACTTGAAA 128456

RESULT 8

AC018145 75650 bp DNA linear HTG 09-DEC-1999

LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.

DEFINITION AC018145

ACCESSION AC018145

VERSION AC018145.1 GI:6553046

KEYWORDS HTG; HTGS_PHASE2.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 75650)

AUTHORS Adams,M. and Venter,J.C.

TITLE Direct Submission

JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10213542 by the submitter. For more information on this record e-mail to fly@celera.com.

* NOTE: This is a "working draft" sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

Location/Qualifiers

1..75650

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

ORIGIN

Alignment Scores:

Pred. No.: 9,12e-298 Length: 75650

Score: 3864.50 Matches: 770

Percent Similarity: 90.69% Conservativeness: 0

Best Local Similarity: 90.69% Mismatches: 4

Query Match: 95.51% Indels: 79

DB: 2 Gaps: 4

US-10-028-384-8 (1-774) x AC018145 (1-75650)

Qy 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20

Db 21487 ATGAATCGAGCGCGAAGATCTGAACAGCAAGTGGCTGCTACAGCAGCCTAATCACC 21428

Qy 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40

Db 21427 TTCGCATCTCTGCTATCGCTGGCTGGCGGATTTCTTCGCTCTTCGCGGTCATC 21368

Qy 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60

Db 21367 CGTTTCGATCGATTCATCATGATTTGATCCGTGGTTCCTAACTACCGGGCCACCGCTAC 21308

Qy 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrPro 80

Db 21307 ATGGTGCAGAAAGGTGTGTACACTTCTCACTGGTTCGACGAGCGCGATGATTCG 21248

Qy 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100

Db 21247 CTCGCGAGGATGTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCGCGGAATC 21188

Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle 109
Db 783 TACCAGGTTGATGATAACAGCTGGCTTATTCATTGGATTAAATACATTGAACATA 842
Qy 110 ProValHisIleArgAspIleCysValPheLeuAlaPheSerGlyLeuThrSer 129
Db 843 ACAGTTCATCAAGAGATGTGTGTATTCCTTGCACCAACTTTTAGCGGCTTACATCC 902
Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAla 149
Db 903 ATATCTAGTTCTCTGCTAACTAGAGAACTCTGGAAACCAAGGACAGACTTCTAGCTGCT 962
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValIleGlySerTyrAspAsn 169
Db 963 TGCCTTCATTTGTCGACAGGATACATATCTCGGTCAAGTGGCGGATCTTGTATTAAT 1022
Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLys 189
Db 1023 GAAGGATTCGCAATTTTGGCTTCAGTTTCATCTACTTATGGTAAAGTCTGTGAAG 1082
Qy 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 1083 ACCGGGTCTGTGTTCTGGACAAATGTGCTGTCTGTCTCATATTTCTACATGTCTCTGG 1142
Qy 210 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle 229
Db 1143 TGGGAGGTTATGTTTCATCATCACTCATCCCTCTCCATGTGTGTGTGTGTGTGTG 1202
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 1203 ATGCAGAGGTACAGCAAGAGAGTCTACATAGCATATAGCATTTCTACATTTGTGGGTTA 1262
Qy 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 1263 ATATATTCATGTCAGATACCTTTTGTGGATTTTCAGCCATCAGAACAGGACGACATG 1322
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 1323 GCAGCTGCAGGTGTCTTTTGGCTGTCTGCAAGCTTACGCTTTTGTGAGTATCTGAGAGAC 1382
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyLeuLeuValGlyVal 309
Db 1383 CGGTGACAAACAGAGATTCAGACCCCTTTCTTTTGGGTCTCTACATAGCTCCAGGC 1442
Qy 310 GlyValPheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
Db 1443 GCTGTGTCTCTAGTGTCTATCTGACATACACAGGTTATATTCACCATGAGTGGC 1502
Qy 330 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 349
Db 1503 AGGTTTATTCATCTAGGATCTGGGTATGCAAAATATACATTCCTCAATTTATTCATCA 1562
Qy 350 ValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuVal 369
Db 1563 GTGCTGAACATCAACGCTACGACATGGGTCTCTTTCTTTCTGATCATATTTCTGTA 1622
Qy 370 CysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1623 TGTACCTTCCGACGAGCCCTATGGTCTGCATCAAAATATCAACGATGAAGAGATATT 1682
Qy 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
Db 1683 GTCCCTCTGTATGGCATCACTGTCTACTTTGCCGAGTGTATGGTGGGCTGTATGCTG 1742
Qy 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal 429
Db 1743 ACTCTGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1802
Qy 430 PheLeuGlnGluAspSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1803 TATTGGGGATGATGAAGAGGAA-----AACCCACCTGTG 1841

Qy 450 AspGluAlaGluAspSerIleGluLysLys-----ThrLeuTyrAspLysAlaGly 466
Db 1842 GAGGACAGCAGTGTAGGATGACAAAAGAAACCCAGGAAACTTTGTATGACAAAGCAGGT 1901
Qy 467 LysLeuLysHisArgThrLysHisAspAlaGlnAspThrGlyValSerSerAsnLeu 486
Db 1902 AAAGTGAAGAAGCATGTGACAGCAAGAAACCTGAAGAGGCTTGGGCCCCAACAATC 1961
Qy 487 LysSerIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThr 506
Db 1962 AAAAGCATTTGACCATGCTGATGCTCATGCTCTCTGTATGATGTTGCGGCTCCATCGCAG 2021
Qy 507 TrpValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGln 526
Db 2022 TGGGTCAACAAGCAACGCTTCTCCAGTCCAAAGTGGTCTTGGCTCTCTCAAT---CAT 2078
Qy 527 AspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsn 546
Db 2079 GATGTGTACAGCAATATATTAGATGATTTTGAAGAAGCTACTTTTGGCTGAGACAAAAC 2138
Qy 547 ThrAlaAspAlaAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMet 566
Db 2139 ACGGATGAACACGCGCGGCTCATGCTGTGGTGGGACTACGCTATCAGATTGCTGGCATG 2198
Qy 567 AlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVal 586
Db 2199 GCCAACAGGACCATCTGTGTGATTAACAACACCTGGAAACAACACGACCATCGCATGTC 2258
Qy 587 GlyLysAlaMetSerSerThrGluLysSerTyrGluIleMetThrSerLeuAspVal 606
Db 2259 GGAAGAGCTATGCTTCCAAATGAACCGCGCTTAAATCAATCATGAGTCCCTTGATGTC 2318
Qy 607 AspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLys 626
Db 2319 GATTATGTGTGTATTTTCGAGAGTGTATGCTGCTATTCGCGGACGATATCAACAAG 2378
Qy 627 PheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAsp 646
Db 2379 TTCTCTGTGATGGTCAGATAGCTGAAGGGAGCATCCCAAGACATCCGGAAGGTGAC 2438
Qy 647 TyrPheThrAspArgGlyGlyPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsn 666
Db 2439 TATTTCCACCCAGCAGGAGAGTTCGAGTGAACAAGCTGGTCTCTCTACTCTGTTAAAC 2498
Qy 667 CysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGly 686
Db 2499 TGCCTTATGATAAAATGTCTACTACAGATTTCGAGAAATGCGAGCTAGATTTTCGCACT 2558
Qy 687 ProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
Db 2559 CCCCCAGGCTTTGACCGAACACAGTAACTGTGAGATTGGAATTAAGACATTAAATTCAG 2618
Qy 707 TyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLys 726
Db 2619 CATTTGGAGAAAGCTTTTACATCAGACGACCTGGCTTGTCAAGATATATAAGTGAAGCA 2678
Qy 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
Db 2679 CCTGAC-----AACAGGAGACACTAGTGCACAAACCTCGAGTCACCAACATCGTCCCC 2732
Qy 744 ProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArg 763
Db 2733 AACAGAAGATATTTCGTAAGAAGACTACTAAAGAGAGCGTGGCTACGTTAAATAAAG 2792
Qy 764 ProValValValLysGlyLysArgThrLeuLys 774
Db 2793 CTAGTGTTTAAGAAAGGCAAGAGACCTCTAAG 2825

RESULT 10

AX799082
LOCUS 2481 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 1 from Patent WO03054008.
ACCESSION AX799082


```

Db 1612 ATTGTCACCATGTTGATGCTGATGCTATGATGATGTTGCTGCTCCACTGATCTGGGTC 1671
Qy ThrSerAsnAlaTyrSerSerProSerLeuValleuAlaPheHisSerGlnAspGly 528
Db 1672 ACAAGCAATGCTACTCTAGTCCAAAGGTAGTCTGCGCCCTCATACAACT--CATGATGCG 1728
Qy 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla 548
Db 1729 ACCAGCAATATCTTAGATGATTTTAGAGAGCTTACTTTTGGCTAAGGCAAAATACAGAT 1788
Qy 549 AspAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
Db 1789 GAACATGACAGGAGTAAATCTCTGTGGGATTAATGGCTATCAGATAGCTGGAATGGCTAAT 1849
Qy 569 ArgTrpThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
Db 1849 AGAATCTAGCTTGGTGATTAATAACACCTGGAATAACAGCCACATAGCAGCTGGTGGAATA 1908
Qy 589 AlaMetSerSerThrGluGlySerTyrGluIleMetThrSerLeuAspValAspTyr 608
Db 1909 GCTATGCTCTTAATGAACAGCAGCCCTATAAATCATGAGGACTCTAGATGATGATTAAT 1968
Qy 609 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
Db 1969 GTTTTGGTATTTTGGAGGGTATTGGCTATTCTGCTGATGATATCAACAAATTTCTC 2028
Qy 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
Db 2029 TGGATGGTATTAGTAGTACGTGAAGGAGAACATCCCAAGAACATTCGGGAAGTGAATTT 2088
Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
Db 2089 ACCCCACAGGAGGAGTTCGGGTAGACAAACAGGATCCCTACTTTGTTGATTCGCTT 2148
Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2149 ATGTATAAATGCTATACATCAGATTTGGAGAAATGCGCTGATTTTCGTACACCCCA 2208
Qy 689 GlyTyrAspArgTrpArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
Db 2209 GGTTTTGGCCGACACGTAATGCTGAGATTTGGAATAAGGACATTAATTCAAACATTTG 2268
Qy 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
Db 2269 GAAGAAGCCTTTACATCAGAACACTGGCTGTGTAGGATATATAAGTAAAGACCTGAT 2328
Qy 729 GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
Db 2329 -----AACAGGAGACATTAAGATCAGAACCTCGAGTCACCAACATTTCCCAAAACAG 2382
Qy 746 AsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
Db 2383 AAGTATTTGTCAAGAGACACTACCAAGGAGGCGTGGCTACATTAATAAATAGCTGGTT 2442
Qy 766 ValValLysGlyLysArgThrLeuLys 774
Db 2443 TTTAAGAAAGGCAAGAAATATCTAAG 2469

```

RESULT 11

```

AY074880
LOCUS 2481 bp mRNA linear PRI 10-DEC-2002
DEFINITION Homo sapiens source of immunodominant MHC-associated peptides
(SIMP) mRNA, complete cds.
ACCESSION AY074880
VERSION AY074880.1 GI:19879588
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2481)
AUTHORS McBride,K., Baron,C., Picard,S., Martin,S., Boismenu,D., Bell,A.,

```

```

Bergeron,J. and Perreault,C.
The model β2m1 minor histocompatibility antigen is encoded by a
mouse homolog of the yeast ST3 gene
Immunogenetics 54 (8), 562-569 (2002)
22326278
PUBMED 12439619
REFERENCE 2 (bases 1 to 2481)
AUTHORS McBride,K. and Perreault,C.
Direct Submission
TITLE Submitted (22-JAN-2002) Molecular Biology Group, Compatisgene, 6100
JOURNAL Royalmount, Montreal, Qc H4p 2R2, Canada

```

FEATURES

```

source
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="3"
1..2481
/gene="SIMP"
1..2481
/gene="SIMP"

```

gene

```

CDS
1..2481

```

```

/notes="similar to Saccharomyces cerevisiae OSTT3 and Mus
musculus protein encoded by GenBank Accession Number
AK018758; source of minor histocompatibility antigen(s)"
/codon_start=1
/product="source of immunodominant MHC-associated
peptides"

```

```

/protein_id="NAL71884.1"
/db_xref="GI:19879589"

```

```

/translation="MAEPSAPESKHSKSLNSPWSGLMALGNSRHGHGFGCAQAKKA
AGAAPKAPAGLSGSLSQPWSLSTFLFLAMAGFSSRLFAVIRFESIHF
DFWYRSTHLSHGFEFLNDFEDRAWYPLGVGTVPVGLMTHLWILNTL
NITVHRDVCFLAFTFSLTISITFLITRELNNQAGLLAAGCTAIAPGVYISRVAG
SPNEGIALFALOFTVYLWKSVKTSVFTMCCLSFYVMVSAWGVYIINLIPLH
VFVLLMORYSKRVYIAYSTFYIYGLIMOI PFVGPQIRTSNHWAAAGVFALQAY
ALQYLRDLTKQBFOTLFFLGVSILAAGAVELSVLYLTCTYIAPNSGRVSLWDYGY
AKIHPIPLASVSEHPTTWSVFFDLHLVCTFPAGLWFCIKNIDRERVVALYISA
VFVAGVPLRMULTPLVVMLSAIAFSNVFHYIGDDMKRENPPVSDSDDRKNGG
NLYDKAGVRGHATCEXTEEGLEGNISIVTMLMLLMFVACHTWVTSNAYSSPS
VVLASVYNDHTNILDREAYFWLRQNTDEHARVNGWDYGVQIAGMARITLVNN
TWNNSHIALVGAAMSNETAAYKIMRTLDVLYIVFGVIGSGDDINKFLWVRIA
EGEHPKDIRSDYFTPGQEFKRVKAGSPTLLNCLMYKMSVYRFGEMOLDPTPPGPR
TNBETGNKDIKHELEAFTSEHMLVRIYKVKAPDNRETLDHKPRVTNIFPKQKYL
KKTTEKRGYIKKLVFKKSKTKTV"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,32e-226 Length: 2481
Score: 2949.00 Matches: 556
Percent Similarity: 83.88% Conservative: 89
Best Local Similarity: 72.30% Mismatches: 112
Query Match: 72.89% Indels: 12
DB: 9 Gaps: 7

```

```

US-10-028-384-8 (1-774) x AY074880 (1-2481)

```

```

Qy 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
Db 187 TCGCAGCGCTGGTGGCGAGTCGCTTCTCTCTCCATCCCTCTTCTCGCTGCTT 246
Qy 30 AlaGlyPheSerSerArgLeuPheAlaValIleAlaArgPheGluSerIleIleGluPhe 49
Db 247 GCCCGGCTTCAGCTCGCGCTCTTCGCCGCTCATCCCTCCGAAAGCATCATCCAGAGTTC 306
Qy 50 AspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
Db 307 GACCCGCTGTTTAACTATAGTCAACACATCATCTTGCACTCTCATGGTCTTATGAATTT 366
Qy 70 LeuAsnTrpPheAsnGluArgAlaTrpTyrProLeuGlyArgIleValGlyThrVal 89
Db 367 TTAATTTGGTTTGAATGAAGAGCATGGTATCCACTAGGAAGATAGTAGGTGGTACTGTT 426
Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle 109

```

Db 427 TACCAGGCTTATGATTAACCGCTGGCTTATTCATTGGATTATTAATCAATTGAACATA 486
Qy 110 ProValHisIleAetAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
Db 487 ACTGTTACATNAAGAGCTATGTGTCTTCAGCACTTTAGCGCCCTTACATCT 546
Qy 130 IleSerThrTyrLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAla 149
Db 547 ATATCTACTTTCCTCTTACAAGAGAACTTTGGAAACCAAGAGCGAGCTTTAGCTGCT 606
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
Db 607 TGTTTTATGTCTATTGTACAGGCTACATAICTGGICAGTAGCTGATCTCTTGATAT 666
Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLys 189
Db 667 GAAGGCATTCCTATTTTGCACCTTCAGTTCACATACTATTATATGGTAAATCTGTAATA 726
Qy 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 727 ACTGGGTCAATTTTGGCAATGTGCTGCTTAICCTATTTCCTATATGGTCTCTGCT 786
Qy 210 TrpGlyGlyTyrValPheIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 229
Db 787 TGGGGTGGTATGTATATCATCAATCTTATTCACCTGCATGCTATTTGTTGTTACTG 846
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 847 ATGCAGATACAGCAAGAGCTTACATAGCATATAGCATTTCTACATTTGGGGTTTA 906
Qy 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 907 ATATTATCAATGCATACCTTTTGTGGGATTCAGCCAAATACAGCAAGAGTGAACATG 966
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 967 GCAGCTGCGAGGTCTTTGCAATGTCTGCAAGCTTATGCTTCTTGGAGTATCTGAGAGAC 1026
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuValGlyVal 309
Db 1027 CGATTAAACAAACAGAGTTCAGACCCCTTTCTTTTGGGTGATCACTAGCTGCAGGT 1086
Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
Db 1087 GCTGTGTTCTTATGCTATCTATTGTGATATATACAGGTTACATITGCACCATGAGTGGC 1146
Qy 330 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaser 349
Db 1147 AGGTTTTATTCATTGTGGGATCTGGGTATGCAAAATATACACATTTCAATATTGTCATCA 1206
Qy 350 ValSerGluHisGlnProThrThrPheSerPhePhePhePhePhePhePhePhePhePhe 369
Db 1207 GTGCTGAGCATCAACCTACGACTTGGGTGCTTCTTCTTGTGATCTACATATTCTTGTA 1266
Qy 370 CysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1267 TGTACCTTCCAGCGCCCTTTGGTCTGCTCAAAATAATCAACGATGAAGAGTATTT 1326
Qy 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
Db 1327 GTTGCTATATGCAATAGTCTGCTACTTTGCTGGAGTGTGGTGGAGTGTGGTGGAGTGTG 1386
Qy 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal 429
Db 1387 ACTTTGACTCCAGTCTGTGTATGTCTGCTGCAATTTGCTTTTCAATGTTTTTGGACAC 1446
Qy 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1447 TATTTGGGGGTGAC---ATGAAAAGGGAAAAATCCACTGTTGGAGCAGCAGGTGATGAG 1503
Qy 450 AspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu--- 468

Db 1504 GAT-----GACAAAAGAAACCAAGGAAATTTGTATGATAAGGCGAGTAAAGTGAGG 1554
Qy 469 LysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSer 488
Db 1555 AAAACATGCACTGACAGGAAAAAACTGAAGAG---GGATTAGCCCTTAATAAAGAGC 1611
Qy 489 IleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpVal 508
Db 1612 ATTTGTCACCATTTGATGCTGATGCTTATGATGATGTTCTGCTGCTGCTGCTGCTGCTGCT 1671
Qy 509 ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
Db 1672 ACAAGCATGCTACTCTAGTCCAAGTGTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1728
Qy 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAla 548
Db 1729 ACCAGGAATATCTTAGATGATTTTAGAGAAGCTTACTTTTGGGTAAAGGCAAAATACAGAT 1788
Qy 549 AspAspAlaArgValMetSerTrpTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsn 568
Db 1789 GAACATGACCGAGTATGCTCTTGTGGGATTTAGCTATCAGATAGCTGGAATGGCTAAT 1848
Qy 569 ArgThrThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
Db 1849 AGAATACGTTGGTGGATTAACACCTGGAATTAACAGCCACACATAGCAGCTGGTGGGAAA 1908
Qy 589 AlaMetSerSerThrGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
Db 1909 GCTATGCTCTTAATGAACAGCAGCCTATAAATCATGAGGACTCTAGATGTAGATTAT 1968
Qy 609 ValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
Db 1969 GTTTTGGTATTTTGGAGGGTTATTGGCTATTCTGCTGATGATATCAACAAATTTCTC 2028
Qy 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
Db 2029 TGGATGTTAGTAGTAGCTGAAGAGAACATCCCAAGACATTCGGGAAGTACTATTTT 2088
Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
Db 2089 ACCCCACAGGAGGAATTCGGTGTAGCAAAAGCAGGATCCCTACTTTTGTGTAATGCTTT 2148
Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2149 ATGTAATAAATGTCTACTACAGATTTGGAGAAATGAGCTGGATTTTCTGACACCCCA 2208
Qy 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
Db 2209 GGTTTTGCCCAACACAGTAAATGCTGAGATTGGAATAAGGACATTAATTTCAACACATTG 2268
Qy 709 GluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHis 728
Db 2269 GAAGAGAGCTTTACATCAGAACACTGGCTTTGAGATATATAAAGTAAAGACCTGAT 2328
Qy 729 GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
Db 2329 -----AACAGGAGACATTTAGATCACAACCTCGAGTCCACACATTTTCCCAAAACAG 2382
Qy 746 AsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
Db 2383 AGATATTGTCAAGAGACTACCAAAAGGAGGCTGCTGCTACATTAATAAATAGCTGTT 2442
Qy 766 ValValLysGlyLysArgThrLeuLys 774
Db 2443 TTTAAGAAAGCAAGAAATATCTAAG 2469

RESULT 12

AX799084 LOCUS 2710 bp mRNA linear PAT 08-OCT-2003
DEFINITION Sequence 3 from Patent WO03054008.
ACCESSION AX799084
VERSION AX799084.1 GI:37605059
KEYWORDS

Qy 607 AspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLys 626
 Db 2030 GATTATGTTGGTTATTTTCGGAGAGTGTATTCGGTATTCGGGGACGATATCAACAAG 2089
 Qy 627 PheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAsp 646
 Db 2090 TTCCTCGATGTCAGATAGCTGAGGGGAGCATCCCAAGACATCCGGGAGGTGAC 2149
 Qy 647 TyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsn 666
 Db 2150 TATTTCACCCAGCAGGAGAGTTCGGAGTAGACAAAGCTGGTCTCTACTCTGTTAAAC 2209
 Qy 667 CysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGly 686
 Db 2210 TGCCTTATGATAAATGCTATCTACATGATTGGAGAAATGACGTAGATTTCGACT 2269
 Qy 687 ProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
 Db 2270 CCCCAGGCTTGAACCAACAGCTAATGCTGAGATTCGAAATAAAGACATTAATCAAG 2329
 Qy 707 TyrLeuGluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLysLys 726
 Db 2330 CATTTGAGGAGAGCTTTTACATCAGACACTGGCTTGTGAGGATATATAAGTGAAGCA 2389
 Qy 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
 Db 2390 CCTGAC-----AACAGGGAGACACTAGGTCACAAACCTCGAGTCACCAACATCGTCC 2443
 Qy 744 ProAlaAsnPhelIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArg 763
 Db 2444 AACAGAGATTTGTTCAAGAGAGACTATAAAGGAAGCGTGGTACGTTAAAAATAAG 2503
 Qy 764 ProValValValLysGlyLysArgThrLeuLys 774
 Db 2504 CTAGTGTTAAGAAAGGCAAGAGACTCTTAAG 2536
 RESULT 13
 UI3019
 LOCUS 54118 bp DNA linear INV 13-JAN-2003
 DEFINITION Caenorhabditis elegans cosmid T12A2, complete sequence.
 ACCESSION UI3019
 VERSION UI3019.1 GI:912482
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 54118)
 Waterston,R.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 99069613
 PUBMED 9851916
 REFERENCE 2 (bases 1 to 54118)
 Latreille,P.
 The sequence of C. elegans cosmid T12A2
 Unpublished (2001)
 REFERENCE 3 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAR-1996) Genome Sequencing Center, Washington
 University
 REFERENCE 4 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUN-1998) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington

University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (28-AUG-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 9 (bases 1 to 54118)
 Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (13-JAN-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 On Jul 27, 1995 this sequence version replaced gi:529354.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: rwenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted; all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T12A2;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C23G10, 200 bp overlap; the 3' cosmid is C18F10,
 15000 bp overlap. Actual start of this cosmid is at base position 1
 of T12A2; actual end is at 39122 of T12A2.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yuji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFeome cloning project (<http://wormfb.dfci.harvard.edu/>),
 similarity to other proteins from BlastX analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans GenBank submissions,
 and personal communications with C. elegans researchers. tRNAs
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
 Location/Qualifiers

FEATURES


```

source
1. 54118
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="III"
/clone="T12A2"
3241. 6926
/genes="hot-4"
/join(3241. 3365,3586. 3749,6392. 6561,6795. 6926)
/genes="hot-4"
/standard_names="T12A2.16b"
/notes="coded for by the following C. elegans cDNAs:
yk1067ell.3, yk1325d06.3"
/codon_start=1
/product="Homolog of odr-2(two) protein 4, isoform b"
/protein_id="AAV15601.1"
/db_xref="GI:20198879"
/translation="MRKWLITVLCTLLQATPFSLYHDSRDKSLMRPKSHCFSCA
SFVYLPNSQLMHVYPPKNFTDCWQPDGIGLVPCCSACFTLVERIDDVEYSEQH
GVIRGMDLLFGDDDVNRLSAYENCRCRHDRLKLLRFLPLSGQTDVTFCSGN
GQFCHEHMLRELSSNQIHSVLLVATWILLF"
/join(3241. 3365,3586. 3740,6392. 6558,6795. 6926)
/genes="hot-4"
/standard_names="T12A2.16a"
/notes="coded for by the following C. elegans cDNAs:
AF324056"
/codon_start=1
/product="Homolog of odr-2(two) protein 4, isoform a"
/protein_id="AAV15600.1"
/db_xref="GI:20198878"
/translation="MRKWLITVLCTLLQATPFSLYHDSRDKSLMRPKSHCFSCA
SFVYLPNSQLMHVYPPKNFTDCWQPDGIGLVPCCSACFTLVERIDDVEYSEQVI
RCMDRLLAGDDDVNRLSAYENCRCRHDRLKLLRFLPLSGQTDVTFCSGDFC
NEHDMRELSSNQIHSVLLVATWILLF"
/complement(7287. 12051)
/genes="T12A2.15"
/complement(join(7287. 7364,7412. 7520,7644. 7735,
7844. 7979,8026. 8132,9511. 9663,9726. 9884,9932. 10111,
10156. 10458,10598. 10687,10735. 10911,10953. 11156,
11304. 11439,11832. 11919,11982. 12051))
/genes="T12A2.15"
/standard_names="T12A2.15a"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
yk194a9.3, yk194a9.5, yk195a9.5, yk818e11.3, yk818e11.5"
/codon_start=1
/product="Hypothetical protein T12A2.15a"
/protein_id="AAC24452.2"
/db_xref="GI:20198877"
/translation="KSMQSYLVPLVGSAILSTFTFLGKWDYFVWVLIILIASVTKS
YLWKEERELIALRATALREVEVIMQAQDLPAWQFQPTDERVEMLNKVIHQLPYVG
EYTKTMDPFIIPVKAQMPGMKFNKFTQMDMDIPCKRVGKIVYTNVGRDRIIVD
MDVAYAGDFTVSCCGFTGGMNIOFSGKRLAILKPLLPVPMVGVSGVTFLEPMKM
DNLTGMGEWELPGLIDAIRSVINSQIALCYLNEIWPVLPADVDVDTQLYFEFPG
VRLKIEAKNLENDISFIKKGSDPFAIOVQSQFRTKRTIDDDLPINWEFEAV
VQADQCKLRIELFDEQGBELGRSLVDLKVQAGRTIDKWPLEGCKRGDLHIKA
TWNLSLELHLEKEWEAEWCAKPIHSHALLMYIDSVADLPYKSKLEFSPFVEV
SLKETQTPVVKVTFVPLFQSKFLFVRLHEGQELKFEAVDDGRRLSGLSINLPLIT
LKPEPLEONQOMHMLTLGVHOSPIVITTRIMRLSDSQSLNSHGRNSRLGLFRS
KHEMKRETRADENGEIRIQIDFDDLVNQLALIRCDLMTDPKQCKNPYYSVKL
VALDGNKVEYKKTPTAKTRHPEDNHEVDINPSDLLNKHVKNVNDTNYGTFFVA
KPVGLCEIRLUSLMRQSRWIFLUSVRK"
/complement(join(7287. 7364,7412. 7520,7644. 7735,
7844. 7979,8026. 8132,8386. 8532))
/genes="T12A2.15"
/standard_names="T12A2.15b"
/notes="contains similarity to Pfam domain PF00168 (C2
domain); coded for by the following C. elegans cDNAs:
yk349e10.5"

gene
1. 54118
/product="Hypothetical protein T12A2.15b"
/protein_id="AAC24451.1"
/db_xref="GI:3258591"
/translation="WNEIVSAFPIERANGKVNQGVVAVPAKTVNQDAADDEVIKLN
VNYOMLRSDQSLSNHSRSLRGLRFLRSHKEMKRETRADENGEIRIQIDFDDLL
VYQKIALIRCDLMTDFKQCKNPYYSVKLVALDGNKVEYKKTPTAKTRHPHFDN
HVEIDINPSDLLNKHVKNVNDTNYGTFFVAKEVPLVGLCEIRLSDLSMRQLSQRMTPLS
VERK"
/complement(16387. 22005)
/genes="T12A2.5"
/complement(join(16387. 16422,16786. 17621,17892. 18452,
18948. 19052,19995. 20008,20190. 20562,20859. 20951,
21039. 22005))
/genes="T12A2.5"
/standard_names="T12A2.5"
/notes="contains similarity to Oryza sativa Putative myosin
heavy chain-like protein.; TR:094DC2; coded for by the
following C. elegans cDNAs: CEES24F"

Alignment Scores:
Pred. No.: 1,74e-195 Length: 54118
Score: 2565.00 Matches: 513
Percent Similarity: 68.38% Conservative: 99
Best Local Similarity: 57.32% Mismatches: 137
Query Match: 63.89% Indels: 147
DB: 3 Gaps: 14

US-10-028-384-8 (1-774) x U13019 (1-54118)
Qy 5 ProLysMetLeuAsnSerLysValAlaGlyTyrSer----- 16
Db 31271 CCAGTCACATATATTCAAACACTCAAGAAATTTCTCTTTTTCAGTAAATGACATCAACAACG 31330
Qy 17 -----SerLeuIleThrPheAlaIleLeu 24
Db 31331 CGCGTCGAAACACGAGTAGTCGCTCGGAGCGACACTCTTCTTACCATCGCTGCTACTG 31390
Qy 25 LeuIleAlaThrLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSer 44
Db 31391 GCTCGCATGTTGCTCGGATTCGCTCCGCTTTCGCCATCGCTCGCTTTCAGTCG 31450
Qy 45 IleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsn 64
Db 31451 ATTATCCAGAGTTCGATCCATCGATTAACTACCTACGCGTACACATCATAGTTTCAACAT 31510
Qy 65 GlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrTyrProLeuGlyArgIle 84
Db 31511 GGTTTCTATAAATTCCTTGAATTTGATGAGCGTGCCTGATCCACTTGGTCGTATT 31570
Qy 85 ValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeu 104
Db 31571 GTCGAGGAACTATATATCCAGACTCATCGGAATCTGGAAATTAATTCACGTGGATTG 31630
Qy 105 HisValLeuAsnIleProValHisIleArgAspIleCysValPheLeuAlaProIlePhe 124
Db 31631 GATTCCCTTAACCTCCATGCTTACATTCGTAAGTTTGGTATTCTTCTGCTCCAAACATTC 31690
Qy 125 SerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAla 144
Db 31691 TCTGGCTGATCCCATTCCTACTTACTTACAAGAGTGTGGAGCCGACGAGCT 31750
Qy 145 GlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArgSerValAla 164
Db 31751 GGCTTTTCCCGCTGCTTCTATGTTATTTCTCCAGGATATACCTCCCGTTCAGTTGCT 31810
Qy 165 GlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrp 184
Db 31811 GGATCTTATGATACGAGGAAATTCATCTTTTCTCTTCAATTCACGTATTATTATCG 31870
Qy 185 ValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPhe 204
Db -----

```

| | | | | |
|-------|----|---|--|-----|
| 479 | QY | ---- | -----AspThrGlyValSerSerAsnLeuIysSerIleValI | 491 |
| 32977 | DB | TCAGAACGCCAACGATGGAGATGACTCGCGAGTTCTCTAATGTCGTCAACATCATTC | 33036 | |
| 491 | QY | eLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTTPValThrSerAs | 511 | |
| 33037 | DB | AATCATCTCTTGTAATCTTCTACTCATGTTCTGTCATGCTACTACGTACATCATCCAA | 33096 | |
| 511 | QY | nAlaTySerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAs | 531 | |
| 33097 | DB | CGCCTATTCTCATCATCTGTCGTCTCTCCAA---AGTTCTACTAACACGAGACCGTAT | 33153 | |
| 531 | QY | nIleLeuAspAspPheArgGluAlaTyTrTyrTriPLeuSerGlnAsnThrAlaAspAsAl | 551 | |
| 33154 | DB | CATCATGGACGATTTCCGTGAGCTTATCATCTGGCTCCGCGAGAACACAGCTGATGATGC | 33213 | |
| 551 | QY | aArgValMetSerTrTPTrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrTh | 571 | |
| 33214 | DB | CGGTGTATTGTCATCGTGGGATTATGGATATCAAAATCGACGAATGGCCAACTACTAC | 33273 | |
| 571 | QY | rLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIysAlaMetSe | 591 | |
| 33274 | DB | ATTAGTTGCACACAAATACCTGGAACAACTCGACATATGCCCTTGTCCGAAAGCCATGTC | 33333 | |
| 591 | QY | rSerThrGluGluIysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuVa | 611 | |
| 33334 | DB | CTCCAAATGAATCAGCGCGTTACGAAATATTACCGCGAACTCGACGTGACTACATCTTGT | 33393 | |
| 611 | QY | lIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnIysPheLeuTrpMetVa | 631 | |
| 33394 | DB | TAATTTCCGAGAGATTATGGAAATCTCGAGATGATATCAACAAATCTTGTGGATGGT | 33453 | |
| 631 | QY | lArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspAr | 651 | |
| 33454 | DB | TCGTATTGCACAAAGAGAACACCCGAAAGATATCCGCGAGGAGAACTATTCCATCGAC | 33513 | |
| 651 | QY | gGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLy | 671 | |
| 33514 | DB | TGTTGAAATATCCACTGGACGAGGTGCTTCTGAAACTATGCTCAATTCGCTCATGTACAA | 33573 | |
| 671 | QY | sLeuSerTyrTyrArgPheGlyGluLeuLeuLeuAspTyrArgGlyProSerGlyTyrAs | 691 | |
| 33574 | DB | AATGACTATTATCGTTTGTGTGAACTCGTGTTGGATACAAACCGAGCTGGTGCTTCGA | 33633 | |
| 691 | QY | pArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGluGluAl | 711 | |
| 33634 | DB | TAGAACTCGTGGATATGTAATTTGAAACAGAGATATCACCTCTGAAATACATAGAAGA | 33693 | |
| 711 | QY | aTyTrThrGluHisTrpLeuValArgIleTyrArgValIysLys---ProHisGluPh | 730 | |
| 33694 | DB | ATATACACGAGAGAACTGGCTTGTGCGTATCTACACGGTAAAGAGCTTCCA----- | 33745 | |
| 730 | QY | eAsnArgProSerLeuIysThrLysGluArgThrIlePro----- | 743 | |
| 33746 | DB | -AACAGACCAACAGTGAAGTGAAGAGGCTACCATCCCAATTAAAGGGGAGAGGCCAC | 33804 | |
| 744 | QY | -----ProAlaAsnPheIleSerArgLysAsnSerLy | 754 | |
| 33805 | DB | ACAGTTACGATTAAAAATATCCACTATTGTGTAATATAAATTAACCATTCGAGGTAA | 33864 | |
| 754 | QY | sArgArgIysGlyTyrIleArgAsnArgProValValIys | 768 | |
| 33865 | DB | AAAAAAAAGGAGCTGATTGCACAGCTCCAAACCGCTTCTAAA | 33907 | |

QY 57 AlaThrAlaTyrMetValGlnAsnGlyTyrPheAsnSerPheLeuAsnThrPheAspGluArg 76
 DB 234 ACAACAAGGTTTCTTCCGCGAAGGATTTATATAATTTCCATAACTGGTTTGGATCGA 293
 QY 77 AlaTyrProLeuGlyArgGlyValGlyThrValTyrProGlyLeuMetIleThr 96
 DB 294 GATGGTATCCATGGGAGAAATATAGGAGGAACAATCTATCCAGGCTCTGATGATACC 353
 QY 97 SerGlyGlyIleHisTrpLeuHisValLeuAsnIleProValHisIleArgAspIle 116
 DB 354 TCTGCAGCATCTACCATATCTACCATTTTCCACATCATCTATTTGACATCCGAATGTT 413
 QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
 DB 414 TGTGATTTCTTGTCTTCCATCTTCTGCTTTTCCACACCATAGTCATCTACATCTGACA 473
 QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaIleSerPheIleAlaIleValPro 156
 DB 474 AAAGAGCTGAAGATGCTGTCGAGCCCTTTGGCAGCAGCAATGATAGCAGTGGTGCCT 533
 QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
 DB 534 GGCATATATCATCTGTCGTCGATCCTATGATGATGAGGATTTGCCATTTCTGT 593
 QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
 DB 594 ATGCTTCTAACTTACTTACATGTTAGGAGTAAAGACTGGTTCAATCTATTTGGCA 653
 QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyValPheIle 216
 DB 654 GCCATGTCGCTCTAGCATATTTCTATATGCTCTGCTGGGAGGATACGTGTTCTG 713
 QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
 DB 714 ATCAACTTATACCATCTTCTGCTGTTTGTGATGCTTACCGAAGATTTCTTCATCGA 773
 QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
 DB 774 ATTATGTCGCTTACTGTCACAGTTTATGTTTGGGCACCATCTCTCCATGCGATCTCC 833
 QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
 DB 834 TTTGTTGGATTCAGGCTGTCTGCTCATCTGACCATGCGCCCTTAGGATTTTGGC 893
 QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 DB 894 CTGTGTCAGATCCATGCTTTGTAGACTACCTCGGAGCAAGCTCAATCTCAGCAATT 953
 QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
 DB 954 GAAATCTTTTAGGAGTGTGATTCACCTGGTTGGATTTGTTCTTCTCATGTCGGA--- 1010
 QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
 DB 1011 -----GCTGCTCATGCTGACC-----GGCAAAATATCCCATGGAAGTCTGCTTC 1058
 QY 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
 DB 1059 TATTCCTCTTTGATCCTTCTTATGCAAAAGCAATATTCATCAATGCTCTGATCT 1118
 QY 352 GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
 DB 1119 GAGCATCAGCCACGACATGCTCTTCTTATTTTGTGATTTGAGTCTTCTTCTCATG 1178
 QY 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
 DB 1179 TTTCCAGTTGGTCTGATATTCCTTCTGAGTAATCTATCTGATGCCAGAAATTTTCATATC 1238
 QY 392 LeuTyrAlaIleSerAlaIleTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
 DB 1239 ATGTACGGTGTGACAGATGATGATCTTTCTGTCAGTCTATGTCGCTGATGCTGTTCTT 1298
 QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431

DB 1299 GCCCTGTTATGTGTATTTCTTCTGCAATGGGTTTCTCAGGTATTTATCCACTTACATG 1358
 QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
 DB 1359 AAGAATCTTGCATAGCGTCTCT----- 1382
 QY 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
 DB 1383 -----GATAAAAACA----- 1394
 QY 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 DB 1395 -----AAAAAGCAACAGGATTTCTACATATCCATCAAAAGCAAGTTGCCAGTGC 1445
 QY 490 ValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysTrpValThr 509
 DB 1446 ATGATCCTGGTCATGGCTTTCTTTTGGTCATTTACATTCATCTCTACATGGGTGACA 1505
 QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 DB 1506 AGTGAAGCTTACTCTCTCTCTCTTATTTGTCGAGCCGCTGGTGA---GATGGTAGC 1562
 QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAsp 549
 DB 1563 AGGATCATCTTGTATGATTTTCAGGGAGGCTTATTTGCTCAGACACACACCCCTGAG 1622
 QY 550 AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
 DB 1623 GATGCAAAAGTAAATGTCGTTGGGATTTATGTTACCATGATACAGCAATGCCAATCGA 1682
 QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
 DB 1683 ACAATCTGCTGATATAATACCTGGAATAACACACACATTTCCAGAGTAGCCGAGCA 1742
 QY 590 MetSerSerThrGluGlyLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
 DB 1743 ATGGCTCCACTGAGAAAGCCATGATGATGATGAGAGAGCTGGATGTTAGTTATGTC 1802
 QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
 DB 1803 CTGTTTATTTTGGGGCTCTCAGGATATTCATCAGATGATATCAACAAATTTTGTGG 1862
 QY 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
 DB 1863 ATGTAAGAAATTCGAGAGCACTGACACTGGTAACAC-----ATCAAGAGCAC 1913
 QY 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
 DB 1914 GATTACTACACTCTCTGAGAGTTTCTGTAGATAGAGAAGGATCTCCCGTCTGCTC 1973
 QY 666 AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
 DB 1974 AACTGCTTATGATACAGATGCTCTATTCGCTTTGGTCAAGTCTACACCGAGCCAA 2033
 QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
 DB 2034 CGCCCTCCAGTTTATGACAGAGTGAGAAACGCAAGAAATCGCAATAAAGATTTCCAGCTT 2093
 QY 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
 DB 2094 GATGTTCTAGAGGAAGCTTACACACAGAGCACTGGCTTGTGAGAATATATAAAGTAAA 2153

Search completed: December 15, 2004, 06:00:01
 Job time : 9328.93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05 ; Search time 892.033 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-8

Perfect score: 4046

Sequence: 1 MNRTPKMLNKGAGYSSLIT.....RRKGYIRNPPVVGKRTLK 774

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131518_8007/app_query.fasta_1.3740
-DB=N Geneseq_23Sep04 -OFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=500 -THR_WAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -WAXLEN=2000000000
-USER=US10028384@cgn2_1.1986 @runat_14122004_131518_8007 -NCPU=6 -ICFU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=1.0 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

N_Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Length | ID | Description |
|------------|--------|--------------|---------|-------------|
| 1 | 4046 | 100.0 | 2417 10 | ADD94789 |
| 2 | 4046 | 100.0 | 2699 4 | ABL13247 |
| 3 | 3918.5 | 96.8 | 4922 4 | ABL13246 |
| 4 | 2949 | 72.9 | 2481 10 | ADD94783 |
| 5 | 2948 | 72.9 | 2710 10 | ADD94785 |
| 6 | 2226 | 55.0 | 2855 4 | ABL02795 |

| | | | | | | |
|----|--------|------|------|----|----------|--------------------|
| 7 | 2198.5 | 54.3 | 3094 | 10 | ADD94791 | ADD94791 Mouse ITM |
| 8 | 2177.5 | 53.8 | 2472 | 10 | ADD94793 | ADD94793 Human ITM |
| 9 | 2177.5 | 53.8 | 2472 | 10 | ADD94793 | ADD94793 Human ITM |
| 10 | 2176 | 53.8 | 2256 | 6 | AB232036 | Ab232036 Candida a |
| 11 | 2172.5 | 53.7 | 2232 | 8 | AB20789 | Ab20789 Aspergill |
| 12 | 2168.5 | 53.6 | 2760 | 5 | ABV24502 | ABV24502 Human pro |
| 13 | 2113.5 | 52.2 | 2466 | 10 | ADB69900 | ADB69900 C. neofor |
| 14 | 2027.5 | 50.1 | 2733 | 10 | ADD94787 | ADD94787 Yeast STT |
| 15 | 2014 | 49.8 | 2603 | 8 | ABT17781 | ABT17781 Aspergill |
| 16 | 2014 | 49.8 | 3969 | 8 | ABT17781 | ABT17781 Aspergill |
| 17 | 2014 | 49.8 | 4603 | 8 | ABT19595 | ABT19595 Aspergill |
| 18 | 1978 | 48.9 | 6153 | 4 | ABL02794 | ABL02794 Drosophil |
| 19 | 1961 | 48.5 | 2000 | 10 | ACC61293 | Acc61293 Gene sequ |
| 20 | 1888.5 | 46.7 | 2000 | 10 | ADK63659 | ADK63659 Disease t |
| 21 | 1888.5 | 46.7 | 1848 | 8 | ABT18969 | ABT18969 Aspergill |
| 22 | 1851.5 | 45.8 | 2882 | 10 | ADB69539 | ADB69539 C. neofor |
| 23 | 1851.5 | 45.8 | 4738 | 10 | ADB69178 | ADB69178 C. neofor |
| 24 | 1840.5 | 45.5 | 1969 | 8 | ABT18375 | ABT18375 Aspergill |
| 25 | 1597 | 39.5 | 1664 | 4 | AAK94164 | AAK94164 Human ful |
| 26 | 1597 | 39.5 | 1664 | 12 | ADL30661 | ADL30661 Full leng |
| 27 | 1512 | 37.4 | 1543 | 2 | AX85055 | AX85055 Human sec |
| 28 | 1512 | 37.4 | 1543 | 8 | ACD18981 | ACD18981 Novel hum |
| 29 | 1512 | 37.4 | 1543 | 12 | ADG78372 | ADG78372 Human sec |
| 30 | 1512 | 37.4 | 1543 | 12 | ADN60663 | ADN60663 Human sec |
| 31 | 1509.5 | 37.3 | 1209 | 6 | AAH33264 | AAH33264 Human col |
| 32 | 1509.5 | 37.3 | 1209 | 6 | ABL89850 | ABL89850 Human col |
| 33 | 1331 | 32.9 | 1371 | 12 | ADP28508 | ADP28508 Human sec |
| 34 | 1191 | 28.4 | 2510 | 5 | AAF93772 | AAF93772 Human cDN |
| 35 | 1191 | 28.4 | 2546 | 2 | AAV44866 | AAV44866 Clone CT5 |
| 36 | 1191 | 28.4 | 2546 | 5 | AAF98463 | AAF98463 Human cDN |
| 37 | 1152.5 | 28.5 | 2537 | 4 | AAO08289 | AAO08289 Human sec |
| 38 | 1147.5 | 28.4 | 2660 | 6 | ABQ54750 | ABQ54750 Human ova |
| 39 | 1147 | 28.3 | 2547 | 4 | AAD08315 | AAD08315 Human sec |
| 40 | 1122 | 27.7 | 2284 | 4 | AAH18021 | AAH18021 Human cDN |
| 41 | 1098 | 27.1 | 2953 | 4 | ABL18224 | ABL18224 Drosophil |
| 42 | 1006.5 | 24.9 | 1114 | 4 | AAH99794 | AAH99794 Human pro |
| 43 | 963 | 23.8 | 787 | 4 | AAH07526 | AAH07526 Human cDN |
| 44 | 923.5 | 22.8 | 764 | 5 | AAF93968 | AAF93968 Primer sp |
| 45 | 851.5 | 21.0 | 2785 | 4 | ABL18208 | ABL18208 Drosophil |

ALIGNMENTS

RESULT 1

ADD94789
ID ADD94789 standard; DNA; 2417 BP.
XX
AC ADD94789;
XX
DT 29-JAN-2004 (first entry)
XX
DE Drosophila melanogaster STT3 gene sequence.
XX
KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide;
KW fruit fly; gene; ds; STT3.
XX
OS Drosophila melanogaster.
XX
FN WO2003054008-A2.
XX
XX 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-CA001967.
XX
PR 20-DEC-2001; 2001US-00028384.
XX
PA (COMP-) COMPATIGENE INC.

| | | | | |
|----|--|----|--|------|
| XX | Perreault C, McBride K; | Db | TCGGGGCGCGCGCTCTTCGCGCGAGCTTCATCGCATCGTGGCTGTACATCAGT | 550 |
| XX | WPI; 2003-559122/52. | Qy | ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaGluPheThr | 180 |
| DR | P-PSDB; ADD94790. | Db | AGTCGCGTGGCTGGATCGTACGATACAGAGGCGATTCGCTATTCGCCCTCGAGTTAC | 610 |
| XX | New human source of immunodominant MHC-associated peptide: (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease. | Qy | LeuPheLeuTyrValArgSerValLysThrGlySerValPheThrSerAlaAlaAla | 200 |
| XX | Claim 6; SEQ ID NO 7; 66pp; English. | Db | TACTTCTGTGGTGGCTCAGTGAAGACTGATCGGTCTCTGTCTGGCGCAGCCGCT | 670 |
| PS | This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC)-associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestinal cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or hematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the Drosophila melanogaster SPT3 gene which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank. | Qy | LeuSerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIle | 220 |
| XX | | Db | TTGTCTCTTCTTACATGTGTTCGCGCTGGGTGTCTACGTGTTCATCAACCTGATA | 730 |
| XX | | Qy | ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer | 240 |
| CC | | Db | CCCTGCAAGCTTCGTACTGTCTATTATGGCAGGTACTCGCCGGTCTGTGACACAGC | 790 |
| CC | | Qy | TyrSerThrPheTyrIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPhe | 260 |
| CC | | Db | TACAGCACCTTCTACATCTCGGACTGCTGTCTCTCCATGCAGATCCCTTCGTGGATTC | 850 |
| CC | | Qy | GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla | 280 |
| CC | | Db | CAACCCGATACGACACAGTGAACATGGCTGGCTGGAGTGTGTGCTCTCTGATGGCC | 910 |
| CC | | Qy | ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe | 300 |
| CC | | Db | GTGGCCACCTTGGCCATTTGGCAGTCCGTGTGTGGCAGACGAGTTCGGAGGTGTTTC | 970 |
| CC | | Qy | IleValGlyLeuLeuValGlyValGlyValPheValAlaValValLeuLeuThrMet | 320 |
| CC | | Db | ATCGTCGGCGGATGTGTGGTGGCTGTGTGGCTGTGTGGCTGTGTGGCTGTGTGGCT | 1030 |
| CC | | Qy | LeuGlyValValAlaAlaProTyrPheSerGlyArgPheTyrSerLeuTyrAspThrGlyTyrAla | 340 |
| CC | | Db | CTGGCGGTGTGGCGCGGTGGAGTGGAGCTTCTACTCTGTGGGATACTGGTACGCG | 1090 |
| CC | | Qy | LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrPheSer | 360 |
| CC | | Db | AAGATCCACATTCCTCATCTTCATCGGTGTGGAGCATCAGCCACCATCTGGTGTCTCG | 1150 |
| CC | | Qy | PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrTyrCysIle | 380 |
| CC | | Db | TTCTTCTTGT | 1210 |
| CC | | Qy | LeuGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe | 400 |
| CC | | Db | AAGCAGATCAACAGCAGCGCTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1270 |
| CC | | Qy | AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly | 420 |
| CC | | Db | GT | 1330 |
| CC | | Qy | ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly | 440 |
| CC | | Db | GTGGCGCTTTTGGGACTGT | 1390 |
| CC | | Qy | ThrAlaIleSerAlaAlaThrGluValAspGluAlaAspSerIleGluLysLysThr | 460 |
| CC | | Db | ACAGCCATAAGCGCAGCCACCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1450 |
| CC | | Qy | LeuTyrAspLysAlaGlyLysLeuLysHisIleArgThrLysHisAspAlaGlnAspThr | 480 |
| CC | | Db | CTGTACGCAAGCTGT | 1510 |
| CC | | Qy | GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMet | 500 |
| CC | | Db | GGCGTCACTTCAACCTTGAAGAGTATGTATTATTGGCCCTTCTAATGTGTGTGTGT | 1570 |
| CC | | Qy | PheAlaValHisCysThrTyrPheValThrSerAsnAlaTyrSerSerProSerIleValLeu | 520 |

Alignment Scores:
 Pred. No.: 0 Length: 2417
 Score: 4046.00 Matches: 774
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-8 (1-774) x ADD94789 (1-2417)

| | | | |
|----|-----|---|-----|
| Qy | 1 | MetAsnArgThrProLysValLeuSerLysValAlaGlyTyrSerSerLeuIleThr | 20 |
| Db | 71 | ATGAATCGGACGCGGAGATGCTGAACAGAGGTGGCTGTGTGTGTGTGTGTGTGTGT | 130 |
| Qy | 21 | PheAlaIleLeuLeuIleAlaTyrPheLeuAlaGlyPheSerSerArgLeuPheAlaValIle | 40 |
| Db | 131 | TTGGCAGTCTGTATTCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG | 190 |
| Qy | 41 | ArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyr | 60 |
| Db | 191 | CGTTTCAGTTCGATTATCCATGATGTTGATCCGTGTGTGTGTGTGTGTGTGTGTGT | 250 |
| Qy | 61 | MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArgAlaTyrPro | 80 |
| Db | 251 | ATGGTGCAGAAATGGT | 310 |
| Qy | 81 | LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle | 100 |
| Db | 311 | CTGGCAGGATTTGGGCGGTACCGTCTATCCCGGCTGTGTGTGTGTGTGTGTGTGTGT | 370 |
| Qy | 101 | HisTyrPheLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu | 120 |
| Db | 371 | CATTGGCTGTGCACGCTACCTCAACATACCGGTCCATTCGTGTGTGTGTGTGTGTGT | 430 |
| Qy | 121 | AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyr | 140 |
| Db | 431 | GGCGCGATCTTCAGTGGCGCTGACCTCCATCTCCACCTACCTGTGTGTGTGTGTGTGT | 490 |
| Qy | 141 | SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer | 160 |

Db 1571 TTCGTGTCCTCACTGCACGCTGGGTGACCAATGCTTACTCCAGTCCCTCCATTGCTTG 1630
QY 521 AlaPheHisAsnSerGlnAspGlySerArgHisLeuLeuAspPheArgGluAlaTyr 540
Db 1631 GCTTTCCCAACAGCTCAGATGATCCCGCACATTTTAGACGATTTTCAGAGAGCTTAC 1690
QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGly 560
Db 1691 TACTGGCTTTCCAGAACACTGCCGATGATGCTGCGTTATGTTCTTGCTGGATACCGA 1750
QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrTrpLeuValAspAsnAsnThrTrpAsnAsn 580
Db 1751 TACCAGATAGCGGGAATGGCAACAGACGCTAGTGGATAATAATACGTGGAACAAT 1810
QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerThrGluGluLysSerTyrGluIle 600
Db 1811 AGTCACATAGCGCTGTTGGCAAGCGCATGCTTCACCGAGAGAGAGTCTACGAAATT 1870
QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
Db 1871 ATGACATCTCTTGACGTGGACTACGTTTGTGTGATCTTTGGCGGTGTGATCGGTATCT 1930
QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
Db 1931 GCGCATGATATCAACAAGTCTCTGTGGATGTCTCGAATTGCTGAGGAGAGCATCCCAAG 1990
QY 641 AspileLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
Db 1991 GACATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGGTAGATCCGAAGT 2050
QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
Db 2051 GCTCCGCGCTGCTCACTGCTTATGTATGATCAAAATTAAAGCTACTACGATTCGGGAATTG 2110
QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
Db 2111 AAGTTGGACTACAGAGTCCATCTGATATGATCGCACAGTAACGCGCTCATTTGGGAAT 2170
QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArg 720
Db 2171 AAGCACTTCGATCTGACCTACCTGAGGAGGCGCTTACACCAAGAACACTGGCTTGTTCG 2230
QY 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
Db 2231 ATCTATAGGTGAAGAGCGCATGAGTTCAATAGACCATCACTGAAGACCAAGGAGAGA 2290
QY 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgArgLysGlyTyrIle 760
Db 2291 ACGATTCTCCAGCAACTTCATTTTCAGAAAGAACTCTAAGCGTCGCAAGGCTACATA 2350
QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
Db 2351 CGAAACCGACCGGTGTGTGTTAAGGGGAAACGAACCTTGAAA 2392

RESULT 2
ABL13247
ID ABL13247 standard; cDNA; 2699 BP.
AC ABL13247;
XX
AC ABL13247;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
OS Drosophila melanogaster.
XX
PN W0200171042-A2.
XX
PD 27-SEP-2001.
XX

PF 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDS; ABB69144.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signaling and cell-cell

XX interactions.

XX Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

XX ABB72072). The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

XX

XX Alignment Scores:

XX Pred. No.: 0 Length: 2699

XX Score: 4046.00 Matches: 774

XX Percent Similarity: 100.00% Conservative: 0

XX Best Local Similarity: 100.00% Mismatches: 0

XX Query Match: 100.00% Indels: 0

XX DB: 4 Gaps: 0

XX US-10-028-384-8 (1-774) x ABL13247 (1-2699)

QY 1 MetAsnArgThrProLysMetIleuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20

Db 80 ATGATTCGAGCGCGGAGATCTGAACAGCAAGTGGTGGCTACAGCAGCTAATCACC 139

QY 21 PheAlaIleLeuLeuIleAlaTyrPheAlaGlyPheSerSerArgLeuPheAlaValIle 40

Db 140 TTGCGCATCTCTGTAATCGCTGGCTGGCCGGATTTTCTCTCGCTCTTCCGCTCATC 199

QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60

Db 200 CGTTTCGAGTCGATTATCCATGAGTTTGTATCCGTTGTTCAACTACCGGCGCCGCTAC 259

QY 61 MetValGlnAsnGlyTyrTyrAsnPheIleAsnTrpPheAspGluArgAlaTyrPro 80

Db 260 ATGTTGCAGAAATGTTGTGTACAACTTCTCACTGAGTTTCGACGAGCGCGATGATCCG 319

QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100

Db 320 CTCGGCAGGATTTGGCGGTACCGTCTATCCCGGCTGATGATTACGTCCGCGGGAATC 379

QY 101 HistTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120

Db 380 CATTTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATCTCGTTCCTG 439

QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140

Db 440 GCGCCGATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGAGCTGTG 499

QY 141 SerAlaGlyIleValGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160

Db 500 TCCGCGGCGCGCGCTCTTTCCGCGCAGCTTTCATGCCATCGTGGCTGCTACATCAGT 559

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyLeuAlaAlaPheAlaLeuGlnPheThr 180
 DB 560 AGGTCGGTGGCTGGATCGTACGATACAGAGGCGATTCGCATATTCGCCCTGCAGTTACCC 619
 QY 181 TyrPheLeuThrValArgSerValLysThrGlySerValPheThrSerAlaAlaAla 200
 DB 620 TACTTCTTGGTGGGCGCTCAGTGAAGACTGGATCCGTTCTGGTCGGCCGACCGCT 679
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleLeuLeuLeu 220
 DB 680 TTGCTCTACTTCTCATGTTGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 739
 QY 221 ProLeuHisValPheValLeuLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 740 CCCCTGACGCTTCTCGTACTGCTCATATGGGAGGTACTCGCCGCTTGTCTGACCCAGC 799
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 800 TACAGACCTTCTCATCTGGGACTGCTGTTCTCCATGCGAGATCCCTTCTGGGATTC 859
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 860 CAACCGATACGACACGATGACATGCTGCGCTGGAGTGTGTGCTCTTATGGCC 919
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 920 GTGGCCACCTTGGCCCATTTGCGAGTCTGCTGTCGCGCAACGAGTTCGGAGAGCTGTT 979
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValLeuThrMet 320
 DB 980 ATGCTGCGCGATTTGCTGGTGGGCTTGGCTGTTGTCGCGCTGCTGCTGCTGCTGCTG 1039
 QY 321 LeuGlyValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 1040 CTGGGCGTGTGGCCCGCTGGAGTGGACGCTTCTACTCGCTGTGGATACCTGGCTACGCC 1099
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 1100 AAGATCCACATTCCTCATGTGATCGTGTGGAGCATCAGCCACCATCTGGTCTCG 1159
 QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 DB 1160 TTTCTTTTGTGTCACATCTGCTGGTGGCTTCCAGTGGAGTGTGGTACTGCTC 1219
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1220 AAGCAGATCAACGACGAGCGCGTTTCTGCTGCTGTACGCCATCAGTGGGTTTACTTC 1279
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1280 GCTGGTGTGATGGTGGCTTGTGATGTTGACCTCACGCCGCTGGTGTGCTGCTGGCGGA 1339
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 1340 GTGGCTTTTGGGACTGTGGATGTGTTCTCTCAAGAGATTCGCTTAAGCGAATGGGC 1399
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460
 DB 1400 ACAGCCATAGCGCAGCAGCCAGGATGGATGAAGCTGAGGATTCATTCAGAGAAGAGC 1459
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnIleAspThr 480
 DB 1460 CTGTACGACAGGCTGGCAGCTGAAGCATGCTGCTAAGCATGATGCCAGCAGGATACT 1519
 QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
 DB 1520 GGGCTCAGCTCCAACTGAAGAGTATGTTATTTTGGCGCTTCTAATGCTGTGTGATG 1579
 QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 DB 1580 TTGCTGTCCACTCAGCTGGTGGACGAGCAATGCTTACTCCAGTCCCTCCATTTGCTGTG 1639
 QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspPheArgGluAlaTyr 540

DB 1640 GCTTTCCACACAGTCAAGATGGATCCCGCAACATTTAGACGATTTTACAGAGGCTTAC 1699
 QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTrpAspTyrGly 560
 DB 1700 TACTGGCTTTCGAGAACACTCCGATGATCTCGGTATGTCTTGGTGGATACCGA 1759
 QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsn 580
 DB 1760 TACCAGATAGCGGAATGCAACAGAACAGACGCTAGTGGGATAATAATACGTGGAACAAT 1819
 QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
 DB 1820 AGTCATAGCTGCTGGTGGCAAGCAATGCTTCAACCGAGGAGAGTCTTACGAAT 1879
 QY 601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620
 DB 1880 ATGACATCTTTGACGCTGGACTACGTTTGTGTGATCTTTGGCGGTGATCGGCTATCT 1939
 QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
 DB 1940 GCGATGATATCAACAAGTCTCTGGATGTTCCGATTTGCTGAGGAGAGATCCCAAG 1999
 QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
 DB 2000 GACATTAAGGAAGCGATTACTTTACCGACCGCGTGAATTCAGGCTAGATGCCGAAGT 2059
 QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680
 DB 2060 GCTCCGCGCTGCTCAACTGCTTATGTACAAATTAAGTACTACAGTTCGGGAATTTG 2119
 QY 681 LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
 DB 2120 AAGTTGGACTACAGAGGCTCATCTGATATGACACACGTAACGCGCTCAITGGGAAT 2179
 QY 701 LysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArg 720
 DB 2180 AAGGATTCGATCTGACCTACCTGGAGGAGGCTACACACAGAACACTGGCTTTGTTCC 2239
 QY 721 IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
 DB 2240 ATCTATAGGCTGAAGAGCGCATGAGTTCATATAGACCATCACTGAAGACCAAGAGAGA 2299
 QY 741 ThrIleProAlaAsnPheIleSerArgLysAsnSerLysArgGlyLysGlyTyrIle 760
 DB 2300 ACATTTCTTCCAGCAAACTTCAATTCGAGAAGAAGAACTCTAAGCGTCCAGGCTACATA 2359
 QY 761 ArgAsnArgProValValLysGlyLysArgThrLeuLys 774
 DB 2360 CGAAACCGACCGGTGTTGTTAAGGGAAGAAAGCAACCTTGAAA 2401

RESULT 3

ABL13246/c

ID ABL13246 standard; cDNA; 4922 BP.

XX AC ABL13246;

XX XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR

PR 11-JUL-2000; 2000US-00614150.
 XX (PEXE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI: 2001-656960/75.
 DR P-PDB; ABB69143.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 34220; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB557737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 0 Length: 4922
 Score: 3918.50 Matches: 772
 Percent Similarity: 90.93% Conservative: 0
 Best Local Similarity: 90.93% Mismatches: 2
 Query Match: 96.85% Indels: 77
 DB: 4 Gaps: 4

US-10-028-384-8 (1-774) x ABL13246 (1-4922)

QY 1 MetAsnArGThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
 DB 3843 ATGAATCGGACCGCCGAAAGATGCTGAACAGCAGAGGTGGCTGGCTACAGCGCTTAATCACC 3784
 QY 21 PheAlaIleLeuLeuIleAlaThrLeuAlaGlyPheSerSerArgLeuPheAlaValIle 40
 DB 3783 TTCGCCATCTCTAATCGCTGGCTGGCGGATTTCTCTCGCTCTTCGCGGTCAATC 3724
 QY 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
 DB 3723 CGTTTCGAGTCGATTATCCATGAGTTTGTATCGGTGTTCAACTACCGGGCCACCGCCTAC 3664
 QY 61 MetValGlnAsnGlyTyrTrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrPro 80
 DB 3663 ATGGTGCAATGGTTGGTACAACTTCCTCAACTGGTTTCGACGAGCGCGATGATCCG 3604
 QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
 DB 3603 CTCGGCAGGATTGTGGGGGTACCGTCTATCCCGCCCTGATGATTACGTCGGCGGAATC 3544
 QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
 DB 3543 CATTCGCTCTGCAGTACTCAACATACCGGTCCATATTCGTGACATCTGCGTGTCTCTG 3484
 QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
 DB 3483 GGGCGCATCTTCAGTGGCCGTACCTCCATCTCCACTACCTGCTGAGCAAGGAGCTGG 3424
 QY 141 SerAlaGlyValAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPrcGlyTyrIleSer 160
 DB 3423 TCCGGGGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCTGGCTACATCAGT 3364
 QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
 DB 3363 AGTCGGTGGTGGATCGTACGATAACGAGGGCATTTGCCATTATTCGCCCTTCAGTTCAAC 3304

QY 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAla 200
 DB 3303 TACTTCTCTGGTGGCTCGCTCAGTGAAGACTGGATCCGTGTTCTGTCGCGCGAGCGCT 3244
 QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
 DB 3243 TTGTCCTACTTCTACATGGTGTGCTCGCTCGGGTGGCTACGTGTTCATCATCACCTGATA 3184
 QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
 DB 3183 CCCCTGACGCTCTTCGTACTGCTCATATGGCAGGTACTCGCGGGCTCTGCTGACACG 3124
 QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPhe 260
 DB 3123 TACAGCACCTTCTACATCTCGGACTGCTGTTCCTCATGCAGATCCCTTCGTGGGATTC 3064
 QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
 DB 3063 CAACCGATACGACGACGATGAAACATGCTGGCTGGAGTGTGTGTCTCTTATGGCC 3004
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
 DB 3003 GTGGCCACTTGGCCACTTTCAGTCTGCTGTGGCGCAACGAGTTCGGAAGCTGTTTC 2944
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 2943 ATCGTCGGCGGATTGCTGTGGCGGTTGGGCTCTTTGTGGCCGCTGCTGCTCACCATG 2884
 QY 321 LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 2883 CTGGCGGTTTGGCCGCTGGAGTGGAGCGTCTTACTCGCTGTGGGATCTATGCTACGCC 2824
 QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 2823 AAGATCCACATTCGCATCATTCGCTGTGCGAGCATCAGCCACCATCTGCTGCTCTCG 2764
 QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrCysIle 380
 DB 2763 TTCTTCTTTGATCTGCACATCTGCTGTGGCGCTTCCACGTGGAGTGTGTACTGCATC 2704
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 2703 AAGCAGATCAACGACGAGCGGTTTTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2644
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 2643 GCTGT 2584
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 2583 GTGGCGCTTTTCGGGACTGTTGGATGTTCTCTGTCAGAGGATTCGCTTAAGCGAATGGGC 2524
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
 DB 2523 ACAGCCATACGCGACGCCACCGAAGTGGATGAGCTGAGGTTCATTTGAGAAAGACG 2464
 QY 461 LeuTyrAspLysAla----- 465
 DB 2463 CTGTACGACAAAGGT- GAGTTCCTTACTAACACATCCATCCGATTTGATTTTAAATACA 2405
 QY 466 -----GlyLysLeuLysHisArgThrIleHisAspAlaGlnGlnAspThr 480
 DB 2404 TCGCATTTGCGAGCTGGCGAGCTGAGCATCTGACTAAGCATGATGCGCCGAGGATACT 2345
 QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuLeuMetMet 500
 DB 2344 GCGGTGAGCTCCAACTCAGAGATATTGTTATTATTGGCCGCTTCTTAATGCTGTGATGATG 2285
 QY 501 PheAlaValHisCysThrTrpValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
 DB 2284 TTCGCTGTTCCTGACGCTGGGTGACGACGCAATGCTACTCCAGTCCCTCCATCTGCTTCG 2225

QY 521 AlaPheHisAsnSerGlnAspGly 528
 DB 2224 GCTTTCACACAGTCAGATGGGTAGACACCAAGTAGCATCATGCTCTCTTTAAACCTA 2165
 QY 529 -----SerArgAsnIleuAspAspPheArgGluAlaTyrTyr 541
 DB 2164 TCTAACCTTTTCTCCACTAGATCCGCAACATTTTAGACGATTTTCAGAGAGCTTACTAC 2105
 QY 542 TrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpTrpAspTyrGlyTyr 561
 DB 2104 TGGCTTTTCGACAGACACTGCGGATGATGCTCGGTTATGCTCTGGTGGATACGGATAC 2045
 QY 562 GlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSer 581
 DB 2044 CAGATAGCGGAAATGGCAACACAGACGACCTAGTGGATAAATAACGTGGAAACATAGT 1985
 QY 582 HistIleAlaLeuValGlyLysAlaMetSerThrGluGluLysSerTyrGluIleMet 601
 DB 1984 CACATAGCGCTGGTGGCAAGCAATGCTTTCACCGAGGAGATCTCTACCAATATAG 1925
 QY 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
 DB 1924 ACATCTCTGACGTGGACTACGTTTGGTGTATCTTTGGCGGTGTGATCGGCTATTCTGGC 1865
 QY 622 AspAspIleAsnLysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAsp 641
 DB 1864 GATGATATCACAGTTCTGTGGATGGTCCGAAATGCTGAGGAGAGATCCCAAGGAC 1805
 QY 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
 DB 1804 ATTAAAGAAAGCGATTACTTTACCGACCGCGGTGAATTCAGGCTAGATCCGAGAGTCT 1745
 QY 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 DB 1744 CCGGCTCTGCTCACTGCTTATGTACAAATTAAGTACTACAGATCGGGAAATGAG 1685
 QY 682 LeuAspTyrArg 685
 DB 1684 TTGACTACAG-GTAAGCGGAAACATTTCTCAGGTAGCGATGCACACTAACTGCCTGT 1626
 QY 686 -----GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPhe 703
 DB 1625 TTCAGAGTTCATCTGATATGATCGCACACGTACCGCTCATTTGGGATTAAGGACTTC 1566
 QY 704 AspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg 723
 DB 1565 GATCTGACCTACCTGGAGGAGGCTACCCACAGAACACTGGGCTGTTCGCACTATATAGG 1506
 QY 724 ValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIlePro 743
 DB 1505 GTGAGAGCGCGATGATTCATATAGACCATCATCTGAAGACCCAGAGAGACGATTCCT 1446
 QY 744 ProAlaAsnPhelleSerArgLys 751
 DB 1445 CCAGCAAACTTCATTTTCAGAAAGGTAGTGTACCTCGACGCTCCCAATGAATTCATTAA 1386
 QY 752 -----AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
 DB 1385 CAATCTGATCTTTGCAGAACTCTAGCGTGGCAAGGCTATACAGAAACCCGCGTT 1326
 QY 766 ValValLysGlyLysArgThrLeuLys 774
 DB 1325 GTTGTAAAGGAAACCAACCTTGAAA 1299

RESULT 4

ADD94783
 ID ADD94783 standard; cDNA; 2481 BP.

XX
 AC ADD94783;

XX
 DT 29-JAN-2004 (first entry)

XX
 DE Human SIMP cDNA sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; anticancer therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW Gene; ss.
 XX Homo sapiens.
 XX WO2003054009-A2.
 XX 03-JUL-2003.
 XX 18-DEC-2002; 2002WO-CA001967.
 XX 20-DEC-2001; 2001US-00028384.
 XX (COMP.) COMPATIGENE INC.
 XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 XX P-PSDB; ADD94784.
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer; or for suppressing an immune response in an autoimmune
 PT disease.
 XX Claim 6; SEQ ID NO 1; 66pp; English.
 XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC anticancer therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.

XX SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,466-291 Length: 2481
 Score: 2949.00 Matches: 556
 Percent Similarity: 83.88% Conservative: 89
 Best Local Similarity: 72.30% Mismatches: 112
 Query Match: 72.89% Indels: 12
 DB: 10 Gaps: 7

US-10-028-384-8 (1-774) x ADD94783 (1-2481)

QY 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuAlaTrpLeu 29
 DB 187 TGCAGCCGCTGGGTGGAGTGGCTTCTCTCTCCATCCATCTCTCTGGCTTGGCTT 246
 QY 30 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleHisGluPhe 49
 DB 247 GCGGCTTCAGCTCGCGCTCTTTCGCGTCATCGCTTCGAAAGCATCATCCAGAGTTC 306

QY 50 AspProTrrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrrTyrAsnPhe 69
 Db 307 GACCCGTTGGTTAACTATAGATCAACATCATCTGCTGCTATGCGGTCTATGAATTT 366
 QY 70 LeuAsnTrrPheAspGluArgAlaTrrTyrProLeuGlyArgIleValGlyThrVal 89
 Db 367 TTAATGTTGGTTGATGAAGAGCATGGTATCCATAGCAAGATAGTAGGTGCTGTT 426
 QY 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrrPheLeuHisValLeuAsnIle 109
 Db 427 TACCAGGGTTGATGATACCGCTGGCTTATTCATGGATTTTAAATACATTTGAACATA 486
 QY 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
 Db 487 ACTGTTCATCAATAGAGAGCATATGATGCTTATTCATGGATTTTAAATACATTTGAACATA 486
 QY 130 IleSerThrTyrLeuLeuThrLysGluLeuTrrPheSerAlaGlyAlaGlyLeuPheAlaAla 149
 Db 547 ATAATCTACTTTCCTGCTTACAGAGAACTTTGGAAACCAAGAGGAGGACTTTTAGCTGCT 606
 QY 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
 Db 607 TGTITTTATTTGTTATGTACAGGCTACATATCTCGGTCACTAGCTGATCTTTGATAT 666
 QY 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrrPheValArgSerValLys 189
 Db 667 GAAGGCATTTGCTATTTTGCACITTCAGTTCACTATCTTATTTATGGGTAAATCTGTAAA 726
 QY 190 ThrGlySerValPheTrrPheAlaAlaLeuSerTyrPheTrrPheValSerAla 209
 Db 727 ACTGGTTCAGTTTTCGACAAATGCTGCTGCTTATCTCTATTTCTATATGCTCTGCT 786
 QY 210 TrrPheGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle 229
 Db 787 TGGGGTGGTTATGATTTATTTATTCATCAATCTTATTTCCACTGCTATGATTTTGTGTTACTG 846
 QY 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTrrPheIleLeuGlyLeu 249
 Db 847 ATGCAGAGATACAGCAAAAGCTCATACAGCATATAGCATTTTACATTTGCGGTGTTA 906
 QY 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
 Db 907 ATATTATCAATGATGATACATCTTTTGGGATTCAGCCAAATCAGAACAGTGAACATG 966
 QY 270 AlaAlaLeuGlyValPheValLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
 Db 967 GCAGCTGCAGGTGCTTTCGATTCGCTGCAAGCTTATGCTTCTGCAAGTATCTGAGAGAC 1026
 QY 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuValGlyVal 309
 Db 1027 CGATTAAACAAAACAGAGTCCAGACCTTTCTTTTGGTCTATCATCTAGCTGCAGGT 1086
 QY 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrrPheSerGly 329
 Db 1087 GCTGTGTTCTTATGTCATCTATTTGATTTATGATTTATGATTTATGATTTATGATTTATG 1146
 QY 330 ArgPheTyrSerLeuTrrPheThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 349
 Db 1147 AGGTTTATTCATTTGCGATATCGGTATGCAAAATACATTTCCAAATTTATGATCA 1206
 QY 350 ValSerGluHisGlnProThrTrrPhePhePhePheAspLeuHisIleLeuVal 369
 Db 1207 GTGCTGAGACATCAACCTACGACTTGGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1266
 QY 370 CysAlaPheProValGlyValTrrTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
 Db 1267 TGTACCTTCCAGAGGCTTTGGTTCTGTCATCAAAATATCAACGATGAAGAGATTT 1326
 QY 390 ValValLeuTyrAlaIleSerAlaValTrrPheAlaGlyValMetValArgLeuMetLeu 409
 Db 1327 GTTGTCTATATCAATCAGTGTCTACTTTTGTGAGTGTATGTTGCGACTGATGTTG 1386
 QY 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal 429

Db 1387 ACTTTGACCTCCAGTGTGTGTGCTGTGCAATTCGCTTTTCAATGCTTTTGGACAC 1446
 QY 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
 Db 1447 TATTTGGGGATGAC---ATGAAGAGGAAATCCACCTGTGGAGACAGCAGTGTATGAG 1503
 QY 450 AspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu--- 468
 Db 1504 GAT-----GACAAAAGAAACCAAGGAAATTTGTATGATAAGGACAGCTAAAGTGGG 1554
 QY 469 LysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSer 488
 Db 1555 AAACATGCACTGAACAGAAAACAACTGAAGAG---GGATTAGGCCCTTAATAAAGC 1611
 QY 489 IleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrrPheVal 508
 Db 1612 ATTGTCCACCATTTGATGCTGATGCTATTTGATGATGATTTGCTGCTCCTGCTGCTGCTG 1671
 QY 509 ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
 Db 1672 ACACGCAATGCTTACTCTAGTCCAAAGTGTAGTCTCGGCCCTCATCAAT---CATGATGGC 1728
 QY 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTrrPheLeuSerGlnAsnThrAla 548
 Db 1729 ACCAGCAATATCTTAGATGATTTTAGAGAGAGCTTACTTTTGGCTAAGCCNAATACAGAT 1788
 QY 549 AspAspAlaArgValMetSerTrrTrrPheTrrPheTrrPheGlnIleAlaGlyMetAlaAsn 568
 Db 1789 GAACATGTCAGCAGTAAATGCTTGTGGGATTTATGCTATATGATAGTGTGGAATGGCTAAT 1848
 QY 569 ArgThrThrLeuValAspAsnAsnThrTrrPheAsnAsnSerHisIleAlaLeuValGlyLys 588
 Db 1849 AGAAGTCTGTTGGTGAATAAACAACCTGGAAATACAGCCACATAGCAGTCTGGTGGGAAA 1908
 QY 589 AlaMetSerSerThrGluGlySerTyrGluIleMetThrSerLeuAspValAspTyr 608
 Db 1909 GCTATGCTTCTTAATGAACACAGCAGCTTATAAATCATGAGGACTCTAGATGTAGATTTAT 1968
 QY 609 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
 Db 1969 GTTTTGGTTATTTTGGGGGTTTATGGCTTATTTCTGCTGATGATATCAACAAATTTCTC 2028
 QY 629 TrrPheValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
 Db 2029 TGGATGTTTAGATAGTGAAGAGAAACATCCCAAGACATTCGGGAAAGTGAATATTTT 2088
 QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
 Db 2089 ACCCACAGGAGAAATTCCTGGTGTAGACAAAGCAGGATCCCTTACTTTTGTGAATTCCTT 2148
 QY 669 MetTyrLysLeuSerTyrTrrPheGlyGlyLeuLysLeuAspTyrArgGlyProSer 688
 Db 2149 ATGTATAAATGTCTACTACAGATTTGGAGAAATGCGCTGGATTTTCTGTACACCCCA 2208
 QY 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
 Db 2209 GGTTTTGCAGCAACACCTAATGCTGAGATTTGGAAATTAAGGACATTAATTTCAAACTTTG 2268
 QY 709 GluGluAlaTyrThrThrGluHisTrrPheValArgIleTrrArgValLysLysProHis 728
 Db 2269 GAAGAAGCCTTTACATCAGAACACTGGCTTTGTAGGATATATAAGATTAAGAACACCTGAT 2328
 QY 729 GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
 Db 2329 -----AACAGGAGACATTTAGATCAAAACCTCGAGTCCCAACATTTTCCCAAAACAG 2382
 QY 746 AsnPheIleSerArgLysAsnSerLysArgGlyGlyTrrPheArgAsnArgProVal 765
 Db 2383 AAGTATTTGTCAAGAGAGACTCCAAAGGAGGCGGTGCTACATTAATAAATAAGCTGGTT 2442
 QY 766 ValValLysGlyLysArgThrLeuLys 774

us-10-028-384-8.rng

Thu Dec 16 16:25:16 2004

2443 TTTRAGAAAGCAAGAAATATCTAAG 2469

DB RESULT 5
ADD94785
ID ADD94785 standard; cDNA; 2710 BP.
XX
AC ADD94785;
XX
DT 29-JAN-2004 (first entry)
XX
DE Mouse SIMP cDNA sequence.
XX
KW source of immunodominant MHC-associated peptide; SIMP; MHC;
KW major histocompatibility complex; human leukocyte antigen; HLA;
KW cystic; immunosuppressive; antisense therapy; gene therapy; cancer;
KW lung cancer; intestine cancer; sarcoma; prostate cancer;
KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
KW haematological cancer; immune response; lymphoid cell proliferation;
KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
KW murine; gene; ss.
XX
OS Mus musculus.
XX
PN WO2003054008-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-CA001967.
XX
PR 20-DEC-2001; 2001US-00028384.
XX
PA (COMP-) COMPATIGENE INC.
XX
PI Perreault C, McBride K;
XX
XX WPI; 2003-559122/52.
XX
DR P-PSDB; ADD94786.
XX
XX New human source of immunodominant MHC-associated peptide, (SIMP) nucleic
XX acids and proteins, useful for diagnosing and treating cancers, e.g. lung
XX or breast cancer, or for suppressing an immune response in an autoimmune
XX disease.
XX
XX Disclosure; SEQ ID NO 3; 66pp; English.
XX
XX This invention relates to a novel isolated or purified human protein,
XX termed source of immunodominant major histocompatibility complex (MHC)-
XX associated peptide (SIMP), which is expressed ubiquitously in human
XX cells, where the protein has the potential of generating several protein
XX fragments binding with high affinity to a human leukocyte antigen (HLA)
XX molecule. The invention may allow development of therapeutics with
XX cytostatic or immunosuppressive activity or provide sequences useful for
XX antisense therapy or gene therapy. The source of immunodominant MHC-
XX associated peptide (SIMP) nucleic acids, proteins and fragments are
XX useful for diagnosing and treating cancers, for example lung cancer,
XX intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
XX cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
XX proteins are also useful for modulating an immune response. Decreasing
XX lymphoid cell proliferation is useful for suppressing an immune response
XX responsible for an autoimmune disease or a transplant rejection. The
XX present sequence is that of the mouse SIMP cDNA which is related to the
XX invention. Note: This sequence does not appear in the specification but
XX was obtained by the indexer from GenBank.
XX
SQ Sequence 2710 BP; 678 A; 648 C; 689 G; 695 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,56e-291 Length: 2710
Score: 2948.00 Matches: 551
Percent Similarity: 83.14% Conservative: 90
Best Local Similarity: 71.47% Mismatches: 114
Query Match: 72.86% Indels: 16
DB: 10 Gaps: 5

US-10-028-384-8 (1-774) x ADD94785 (1-2710)
QY 10 SerIysValAlaGlyTyrSerSerIleuThrPheAlaIleLeuLeuLeuAlaTrpIeu 29
DB 254 TCGCAGCGCGCGGTCGTCGTCCTTCCATCCATCCCTCTCCCTGGCGCTGGCTG 313
QY 30 AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
DB 314 GCCGGCTTCAGCTCGCGCTCTTGGCGCTCATCGCTTCGAGAGCATCATCACGAGTTC 373
QY 50 AspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
DB 374 GACCCGCTGGTTTAACTATAGATCAACACATCATCTTGCATCTCTATGGATCTTATGAGTTT 433
QY 70 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrVal 89
DB 434 CTAAATTGGTTTGTATGAAGAGCATGGTACCCACATCGGGAAGATAGTGGTGGCACCGTT 493
QY 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValIleuAsnIle 109
DB 494 TACCCAGGGTTTCATGATTAACAGCTGGCTTATTCATTGGATTTTAAATACATTCAGACATA 553
QY 110 ProValHisIleArgAspIleCysValPheIleuAlaProllePheSerGlyLeuThrSer 129
DB 554 ACAGTTTCACATAAGAGATGTGTGTATTTCTTGTCCCTTGTACCAACCTTTTAGCGGCTTACATCC 613
QY 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAla 149
DB 614 ATATCTACGTTCTCTTACTAGAGACTGTGTGAACCAAGAGAGAGAGAGAGAGAGAGAGAG 673
QY 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
DB 674 TGCTTCATTGCTATCGTACAGGGTACATATCTCGTCCAGTGGCGGGATCTCTTGTATAT 733
QY 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValIys 189
DB 734 GAAGGATTTGCCATTTTGGGCTTCAGTTCACTTACTTACTTATGGTAAAGTCTGTGAAG 793
QY 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
DB 794 ACCGGGCTGTGTCTCGACAATGTCTGCTGTCTATATTTCTACATGTGCTCTCTGGG 853
QY 210 TrpGlyGlyTyrValPheIleIleAlaLeuLeuIleProLeuHisValPheValLeuLeuIle 229
DB 854 TGGGAGGTTATGTGTTCATCATCAACCTCACTCCCTCTCATGTGTGTGTGTGTGTGTGTGT 913
QY 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerTyrPheTyrIleLeuGlyLeu 249
DB 914 ATGCAGAGGTACAGCAAGAGAGTCTACATAGCATATAGCATTTGTACATTTGTGGTTTA 973
QY 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
DB 974 ATATTATCCATGAGATACCTTTTGGGATTTTCCGCAATCGAACAAGAGAGAGAGAGAG 1033
QY 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisIleuGlnSer 289
DB 1034 GCAGCTGCAGGTGTCTTGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1093
QY 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyVal 309
DB 1094 CGGTTGACAAACAGGAGTTCCAGACCCCTTTCTTTTGGGTGCTCTCATAGCTGAGGAGGC 1153
QY 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
DB 1154 GCTGTGTCTTGT 1213
QY 330 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaIleHisIleIleIleIleIleIleIle 349
DB 1214 AGGTTTATTCATCTATGGGATCTGGGTATGCAAAAATACACATTCATTCATTCATTCATTC 1273
QY 350 ValSerGluHisGlnProThrThrTrpPheSerPhePhePheAspLeuHisIleLeuVal 369

Db 1274 GTGTCGACATACACGCTACGACATGGGTGCTTCTTCTTCTTGTATCTACATATCTTGTGA 1333
QY CysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1334 TGTACTCTCCAGCAGGCTATGTTCTGCATCAAAAATATCAACGATGAAGATATTT 1393
QY 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
Db 1394 GTCGCTCTGTATGCGATCAGTGTGCTGTTCTCCGAGGTATGTTGCGGCTGTATGCTG 1453
QY 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspVal 429
Db 1454 ACTCTGACCCCGTCTGTCGATGCTGTCGCCATCGCTTCTCCATGTTTGTGACAC 1513
QY 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1514 TATTGGGGGATGATCAAAAGGAA-----AACCCACCTGTG 1552
QY 450 AspGluAlaGluAspSerIleGluLysLys-----ThrLeuTyrAspLysAlaGly 466
Db 1553 GAGGACAGCTGATGAGGATGACAAAGAAACCCAGGAACTTGTATGACAAAGGAGGT 1612
QY 467 LysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeu 486
Db 1613 AAGTGAGGAACATGTGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1672
QY 487 LysSerIleValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThr 506
Db 1673 AAAGCATTTGACCATGCTGATGCTCATGCTGCTGATGTTTCCGGGTCCACTGCACG 1732
QY 507 TTPValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGln 526
Db 1733 TGGGTACAGCAACGCTTACCTCCAGTCCAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1789
QY 527 AspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsn 546
Db 1790 GATGTTACCAAGATATATTAGATGTTTATGAGAGCGTACTTTTGGCTGAGACAAAC 1849
QY 547 ThrAlaAspAlaArgValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMet 566
Db 1850 ACGGATGAACACGCGGCTCATGTCGGTGGGACTACGGCTATCAGATGCTGGCATG 1909
QY 567 AlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuVal 586
Db 1910 GCCAACAGGACCATCTGTTGGATTAACCAACCTGGAAACACGACCATCGCATGCTGC 1969
QY 587 GlyLysAlaMetSerSerThrGluLysSerTyrGluIleMetThrSerLeuAspVal 606
Db 1970 GGAAGAAGCTATGCTTCCAAATGAACGCGCGCTATAAATCATGAGGTCCTCTGATGC 2029
QY 607 AspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLys 626
Db 2030 GATTATGCTGTTGTTATTTTCGAGAGTATGCTATTTCCGGGGACGATATCAACAAG 2089
QY 627 PheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAsp 646
Db 2090 TTCCTCTGATGGTCAGGATACCTGAAGGGGAGATCCCAAGACATCCGGGAAGGTGAC 2149
QY 647 TyrPheThrAspArgGlyGluPheArgValAspAlaGlyAlaPheAlaLeuLeuAsn 666
Db 2150 TATTTTCAACCCAGCAGGAGGTTCGAGTAGACAAAGCTGGTCTCTCTACTCTGTTAAAC 2209
QY 667 CysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGly 686
Db 2210 TGCTTATGTATAAATGTCATCTACAGATTTGGAGAAATGCGAGTATGTTTCCGCACT 2269
QY 687 ProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
Db 2270 CCCCCAGGCTTTGACCGACACGTAATGCTGAGATTGGAATTAAGACATTAATTTCAAG 2329
QY 707 TyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLys 726
Db 2330 CATTTGGAGGAGCTTTTACATCAGACGACCTGGCTTGTGAGGATATATAAGTGAAGCA 2389

QY 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
Db 2390 CCTGAC-----AACAGGAGACACTAGGTCAAAAACCTCGAGTCACCAACATCGTCCC 2443
QY 744 ProAlaAsnPheIleSerArgLysAsnSerLysArgGlyTyrIleArgAsnArg 763
Db 2444 AACAGAGATATTTGTCAAGAGACTACTATAAAGGAGCGTGGCTACGTTAAATAAATAG 2503
QY 764 ProValValValLysGlyLysArgThrLeuLys 774
Db 2504 CTAGTGTTTAAGAAAGCAAGACACCTCTAAG 2536

RESULT 6

ABL02795

ID ABL02795 standard; cDNA; 2855 BP.

XX

AC ABL02795;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW Pharmaceutical; gene; ss.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

XX

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656960/75.

XX

DR P-PSDB; ABB58692.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

XX

PS Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU1840-ABU16175) and the encoded proteins (ABB57737-CC ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3,58e-217 Length: 2855
Score: 2226.00 Matches: 417
Percent Similarity: 74.20% Conservative: 115
Best Local Similarity: 58.16% Mismatches: 145
Query Match: 55.02% Indels: 40
DB: 4 Gaps: 8

US-10-028-384-8 (1-774) x ABL02795 (1-2855)

QY 18 LeuIleThrPheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerSerArgLeuPhe 37
DB 467 CTGTCAGAGTGGCCATCTCACTCCGCGAGCGGTATTATCAATTTGCCACACGCTGTTC 526
QY 38 AlaValIleArgPheGluSerIleIleHisGluPheAspTrpPheAsnTrpArgAla 57
DB 527 TCTGTGTCGATTCGAAGCGTAATCAATGATCCGATCCGATCTTCAACTACCGCACC 586
QY 58 ThrAlaTrpMetValGlnAnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAla 77
DB 587 ACGCGGTTCGCGGAGCAGGCGCTTTTCAAGTTCACACATGTTTCGATACCGCCGC 646
QY 78 TrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSer 97
DB 647 TGGTATCCCTGGCGCCGATCATCGCGCGGCACCATCTATCCCGGCGTATGCTCACTCG 706
QY 98 GlyGlyIleHisTrpLeuLeuHisValLeuAnIleProValHisIleArgAspIleCys 117
DB 707 CGCGCCCTGACCGCTGATGTGCTCAATGTGACCATGACATGACATGACGAACTGTGC 766
QY 118 ValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIlys 137
DB 767 GTCTTCCTGGCGCCCTTCTCTCTGCTGACACACGCTGTGACCTACGCCCCTCAAAAG 826
QY 138 GluLeuTrpSerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValProGly 157
DB 827 GAGATACACAGCACTGAGCTGAGCTGGTGGCGCGCGCTTTCATATCCATCGTTCCCGG 886
QY 158 TyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeu 177
DB 887 TATATCTCTCGATCCGTGGCGGATGTCACGAATGAAGGATCGCCATTTTCTGCATG 946
QY 178 GlnPheThrTyrPheLeuTrpValArgSerValIlyThrGlySerValPheTrpSerAla 197
DB 947 CTCTTCACCTACTATTGTGGATCAAGCGTGAAGACCGGACGATCTTTTGTGCGCT 1006
QY 198 AlaAlaLeuSerTyrPheTyrMetValSerAlaTrpIleGlyTyrValPheIleIle 217
DB 1007 ATGTCGGCATTTGGCTCTATATATGCTCTCTCGTGGGTGGCTATGCTCTCTGAT 1066
QY 218 AsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeu 237
DB 1067 AACCTATATCCGCTGACGCTGGCGCTGATGATCACCGGACGTTTCTCGACAGATC 1126
QY 238 LeuThrSerTyrThrPheTyrIleLeuGlyLeuLeuPheSerMetClnIleProPhe 257
DB 1127 TACATAGCATACACAGCTATATCTGCTCGGCACCATTTCTGTCGATGACGATCTCGT 1186
QY 258 ValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeu 277
DB 1187 GTGGGATTCACACCATCCAGACTCCGACATCCGACATCTGCTGGAACCTTTGCGCTG 1246
QY 278 LeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArg 297
DB 1247 TGGCAGATTCACGCTTTCTGTCGATCTGCTGCGCATTCCTCCCAAGGATCACCTCGAT 1306
QY 298 LysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaVal 315
DB 1307 CTGCTCTTCAAGAGTGGTTCAGTGTTCG-----ACTGTGCTGTCTGCTGGGT 1360
QY 316 ValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrp 335
DB 1361 ACCCTGTCTACGCTTACCGGGAAGTCTCCCTCGGACCGGAGATTTTACTCGCTACTA 1420
QY 336 AspThrGlyTyrAlaIlyIleHisIleProIleAlaSerValSerGluHisGlnPro 355
DB 1421 GATCATCTTATGCGAGATCATCATCCATCATCCCTCGTATCGGAGATCAGCCT 1480
QY 356 ThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAlaPheProValGly 375
DB 1481 ACATCTCGGTGCTTTCTATTATGATCTGACATCTCTGCTCTCTCTCTCTCTCTCTCT 1540
QY 376 ValTrpTyrCysIleIlysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIle 395

DB 1541 CTTTATTTCTGCTTTCCAGCTGACGACTCGAACATCTTTATTTATTTTGTATGGCGTC 1600
QY 396 SerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValVal 415
DB 1601 ACCAGTATTATTTCGCGGAGTGTGCTGCTTAATGCTGGCTCGCGCCCGTGTATG 1660
QY 416 CysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGlu----- 433
DB 1661 TGTGTGCTATCCGGAATGCCATTTCCGATCTGCTGCGCAAGTATATCAAGAGCGTTGAT 1720
QY 434 ---AspSerSerIlyArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAla 452
DB 1721 GCGGAGAGCTCATGAAGCAGGGA----- 1744
QY 453 GluAspSerIleGluIlyValSerLeuTyrAspIlyAlaGlyLysLeuLysHisArgThr 472
DB 1745 -----GTAGAAGCAGCGCGCAGCAGCAAGAGCTG----- 1774
QY 473 LysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeu 492
DB 1775 -----GAACACGACGACGCGGAGCGTGAAGACCGAGGTGGCCATTTGATTCGTCGGT 1825
QY 493 AlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAla 512
DB 1826 GTCATCATCTCATGCTTATAGTGTACACGCTGCACCTGCACCTGGTCACTCGAGGCC 1885
QY 513 TyrSerSerProSerIleValLeuAlaPheHisSerGlnAspGlySerArgAsnIle 532
DB 1886 TACTCTCTCCCGAGTATTGTGTTGAGT---OCCAGGTGCGACGATGGCGCCGCTCAT 1942
QY 533 LeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArg 552
DB 1943 TTCGATGACTTCGCGAGGCGCTACTACTGCTGCGAGTGAACATCTCGGAGGACGCTCGC 2002
QY 553 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 572
DB 2003 ATAATGTCCTGTGGGATCTACGCTACCATTAACGCGCATGCGCAATCGGACGATATG 2062
QY 573 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
DB 2063 GTGGATAACAATACTTGTGAACACACACATATATCGCGCTGGCGAGCGATGGCTCT 2122
QY 593 ThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
DB 2123 TCGGAGGAGAAAGCCTACGAGATATGAGGAACCTGGATGCGACTACGTTCTCGTGAT 2182
QY 613 PheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrpMetValArg 632
DB 2183 TTCGAGGCGCTCACTGGCTACTCATCGACGATATCAACAGTTCTCTGTGGATGGCGC 2242
QY 633 IleAla-----GluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 2243 ATTGCGCGCAGCACCGATCTGTGTGCGCAC-----ATCCGCGAAAGGACTACTAT 2293
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
DB 2294 CGGCGCAACGAGAGTTCGAGTGGACAAGGAGGCTCACCCACACTGCTCAATTGTTG 2353
QY 669 MetTyrLysLeuSerTyrTyrArgPheGlyLeuLeuLysLeuAspTyrArgGlyProSer 688
DB 2354 ATGTACAAGATGTGCTACTATCTGCTTTGGGCAATATACCGAAGGTGGCAAGCCGAG 2413
QY 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
DB 2414 GGTACATCGAGTTCGTCGCGCGAGATCGCAACAGGACTTTGAATGGATGTATG 2473
QY 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLys 725
DB 2474 GAGGAGCGCTACCCACCGAGCAGCTGGCTGTGTCGATCTACAAAGGTTAAG 2524
RESULT 7
ADD94791

Db 1222 TTCCAGTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCTGGATTTTATCATC 1281
 Qy 392 LeuTyAlaIleSerAlaValTyPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
 Db 1282 ATGTATGGTGTGACCATGATGCTTTTACGCTGTAATGCTGCTGCTGCTGCTGCTG 1341
 Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuValPheLeu 431
 Db 1342 GCACCTGTTATGTCATCTTTTCCGAGTGGTGTTCCTCCAGTGTCTGTCACATATG 1401
 Qy 432 GluGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
 Db 1402 AAAAATCTGCATAGTCCCA----- 1425
 Qy 452 AlaGluAspSerLeuGlyLysThrLeuTyAspLysAlaGlyLysLeuLysHisArg 471
 Db 1426 -----GACAGAGAGCAAG----- 1440
 Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 Db 1441 -----AAGCACAGGATTCTACTCCCTATTAAAGATGAGTGGCGAGTGG 1488
 Qy 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
 Db 1489 ATGATCTGTCATGCTGCTTTTCTCATCCTACCTGCTTTTCTCATGCTGGTGACC 1548
 Qy 510 SerAsnAlaTySerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 Db 1549 AGTGAAGCTTATCTCTCTCATGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1605
 Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTyTrpLeuSerGlnAsnThrAlaAsp 549
 Db 1606 AGGATCATTTTGTATGCTTCCGAGAGCGGTATTATGCTGCTGCTGCTGCTGCTGCTG 1665
 Qy 550 AspAlaArgValMetSerTrpTyrAspTyGlyTrpGlnIleAlaGlyMetAlaAsnArg 569
 Db 1666 GATGCAAAAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1725
 Qy 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
 Db 1726 ACAATTTTGTGACATATACATGCAATATATACCATATTTCTCGATGAGGCGAGCA 1785
 Qy 590 MetSerSerThrGluGlyLysSerTyGluIleMetThrSerLeuAspValAspTyVal 609
 Db 1786 ATGCAATCCACAGAGAAAGGCTATGAAATCATGAGGAGCTTGATGCTGCTGCTGCTG 1845
 Qy 610 LeuValIlePheGlyGlyValIleGlyTySerGlyAspAspIleAsnLysPheLeuTrp 629
 Db 1846 CTGTGCTATTTTGGAGCCCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1905
 Qy 630 MetValArgIle---AlaGluGlyGlyHisProLysAspIleLysGlySerAspTyPhe 648
 Db 1906 ATGCTCGGATGAGGAGAACACAGACAGAGAGAGACATTAAGAGAAATGACTACTAT 1965
 Qy 649 ThrAspArgGlyGluPheArgValAspAlaGlyAlaProAlaLeuLeuAsnCysLeu 668
 Db 1966 ACTCCTACTGGGAATTCCTGTTGATGCTGAGGGTTCCTCGGCTGCTGCTCAACTGCTT 2025
 Qy 669 MetTyLysLeuSerTyTrpArgPheGlyGluLeuLysLeuAspTyArgGlyProSer 688
 Db 2026 ATGTACAAATGCTTACTACCTGTTTGGCAGGCTCTACACAGAGCCAGGCTCCACCA 2085
 Qy 689 GlyTyAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyLeu 708
 Db 2086 GGCTTTGACCGGTTCGAAATGCTGAGTTGGTAAATAAGACTTTGAGCTTGATGCTGCTG 2145
 Qy 709 GluGluAlaTyThrThrGluHisTrpLeuValArgIleTyArgValLys 725
 Db 2146 GAGGAAGCGTATACCAACAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2196

XX ADD94793;
 AC 29-JAN-2004 (first entry)
 XX Human ITM1 gene sequence.
 DE
 XX source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ds; ITM1.
 XX Homo sapiens.
 OS
 XX WO2003054008-A2.
 PN
 XX 03-JUL-2003.
 PD
 XX 18-DEC-2002; 2002WO-CA001967.
 PF
 XX 20-DEC-2001; 2001US-00028384.
 PR
 XX (COMP-) COMPATIGENE INC.
 PA
 XX Perreault C, McBride K;
 PI
 XX WPI; 2003-559122/52.
 DR
 XX P-PSDB; ADD94794.
 DR
 XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.
 XX
 PS Disclosure; SEQ ID NO 11; 66pp; English.
 CC
 CC This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcomas, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human ITM1 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.
 XX
 SQ Sequence 2472 BP; 566 A; 569 C; 583 G; 755 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,69e-212 Length: 2472
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 53.82% Indels: 51
 DB: 10 Gaps: 9
 US-10-028-384-8 (1-774) x ADD94793 (1-2472)
 Qy 17 SerLeuIleThrPheAlaIleLeuLeuAlaTrpLeuAlaGlyPheSerArgLeu 36
 ID ADD94793 standard; DNA; 2472 BP.

RESULT 8
 ADD94793
 ID ADD94793 standard; DNA; 2472 BP.

| | | | |
|----|------|---|------|
| Db | 152 | ACACITTTGAAGCTTCTCATTCTGTCGAATGGCTGCTGTATTATTCCTTCTCCACCTGCTG | 211 |
| Qy | 37 | PheAlaValIleAArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg | 56 |
| Db | 212 | TTTGTCGTCCTGAGATTGAAAGGTATTATCCATGAGTTTGATCCGTACTTTTAATATCGG | 271 |
| Qy | 57 | AlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArg | 76 |
| Db | 272 | ACTACACAGTTCTCGCTGAGAGGGGTTTATAAATCCATAATCGTTGTGATGACCGGA | 331 |
| Qy | 77 | AlaTrpTyrProLeuGlyAArgIleValGlyThrValTyrProGlyLeuMetIleThr | 96 |
| Db | 332 | GCCTGCTACCTTTGGAGCAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATCAC | 391 |
| Qy | 97 | SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleActAspIle | 116 |
| Db | 392 | TCTGTCGAATCTACCATGATCTCCATTTTCCACATCACCATTCGATTCGGAAATGTC | 451 |
| Qy | 117 | CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr | 136 |
| Db | 452 | TGTGTGTTCTCGGCCCTCTCTCTCTCTCCATTCACCTCCATCGTCACGTACTCTCTTAC | 511 |
| Qy | 137 | LysGluLeuTyrSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro | 156 |
| Db | 512 | AAAGAGCTCAAGGATGACAGGGCTGGGCTTCCTGCTGCCATGATGTCGTGATTCCT | 571 |
| Qy | 157 | GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla | 176 |
| Db | 572 | GGATATATCTCCCATCTGTGCTGGCTCTCATATGTAATGAAGGATTTGCCATCTTTTGC | 631 |
| Qy | 177 | LeuGlnPheThrTyrPheLeuTyrValArgSerValIysThrGlySerValPheTrpSer | 196 |
| Db | 632 | ATGCTACTCACCTACTACATGTCGATCAAGGCAAGTAAGACCTGTTCCATCTGTGGGCA | 691 |
| Qy | 197 | AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle | 216 |
| Db | 692 | GCTAAGTGTGGCCCTGTCTATTCTACATGGTCTCTCATGGGAGGTATGTCTTCTCG | 751 |
| Qy | 217 | IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlyArgTyrSerProArg | 236 |
| Db | 752 | ATCAACTTAATTCCTCTCCAGTCTCTGTCGTCGATGCTCACAGCCGTTTCTCTCACCG | 811 |
| Qy | 237 | LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro | 256 |
| Db | 812 | ATCTATGTGGCCTACTGTACTGTTTACTGCTCGGTACTATATCTTTTAGCGCATCTCC | 871 |
| Qy | 257 | PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal | 276 |
| Db | 872 | TTTGTGGTTTCCAGCCCTGCTTTTCATCAGACACATGTCAGCGGTTTGGGCTCTTTGGT | 931 |
| Qy | 277 | LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe | 296 |
| Db | 932 | CTCTGCCAGATCCATGCCCTTTGGATTACTCTCGCAGCAAGTTGAATCCACAACATTT | 991 |
| Qy | 297 | ArgLysLeuPhe-----IleValGlyLeuLeuValGlyValGlyVal | 311 |
| Db | 992 | GAAGTTCCTTTCCGGAGCGTCATCTCTGCTGATAGGCTTTGCTCTCTCACCGTGGGA-- | 1048 |
| Qy | 312 | PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe | 331 |
| Db | 1049 | -----GCTCTCTCATGCTGACA-----GAAAAAATATCTCCCTGACGGGGCGTTTC | 1096 |
| Qy | 332 | TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSer | 351 |
| Db | 1097 | TACTCACTGCTGGATCCCTCTATGCTTAAGAAACAACATCCCCATCATTTGCTTCTGTCT | 1156 |
| Qy | 352 | GluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla | 371 |
| Db | 1157 | GAGCATCAGCCCAACACTGCTCTCATACTATTTTGACTCTGACCTCTCTGTTTCATG | 1216 |
| Qy | 372 | PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal | 391 |
| Db | 1217 | TTTCAGTTGGCCTATTATCTGTTTAGCAACTGCTGATGCCCGGATTTTATCATC | 1276 |

RESULT 9
ADH28844
ID ADH
XX

QY 472 ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 DB 1436 -----AACCAACAGATTCACCTACCTATTAAGATTGAAGTGGCAAGTGGG 1483
 QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
 DB 1484 ATGATACTGGTCATGGCTTCTTCTCATCACTACACACCTTTCATTCAACCTGGTGACC 1543
 QY 510 SerAsnAlaValSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 DB 1544 AGTGAAGCTACTCTTCTCCCTCCATGTACTATCTGCCCTGGTGGG---GATGGCAGT 1600
 QY 530 ArgAsnIleLeuAspPheArgGluAlaValTrpLeuSerGlnAsnThrAlaAsp 549
 DB 1601 AGGATCATATTTGATGACTTCGAGAGCAATATTATTGGCTTCGTCTATACTACCTACAG 1660
 QY 550 AspAlaArgValMetSerTrpAspTyrGlnIleAlaGlyMetAlaAsnArg 569
 DB 1661 GATGGAGGTCATGCTCTGGTGGATATGCTATCATGATTCAGTATGGCAACCGA 1720
 QY 570 ThrThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
 DB 1721 ACAATTTAGTGGACAATAACACATGGAATAATACCATATTTCTCGAGTAGGGCGCA 1780
 QY 590 MetSerSerThrGluGluLysSerTyrGlnIleMetThrSerLeuAspValAspTyrVal 609
 DB 1781 ATGGCGTCCACAGAGGAAAGGCTATGAGATCATGAGGAGCTCGATGTACGATATGTG 1840
 QY 610 LeuValIlePheGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp 629
 DB 1841 CTGGTCATTTTGGAGGCTCCTCTGGTATTCCTCTGATGATATCAACAAGTTCTTTGG 1900
 QY 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
 DB 1901 ATGGTCCGATTTGGAGGAGCACATACAGCAACAT-----ATCAGGAGAAAT 1951
 QY 646 AspTyrPheThrAspArgGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
 DB 1952 GACTATTATCTCCCAACTGGGAGTTCGGTGGACCGTGAAGTTCTCCAGTGCTGCTC 2011
 QY 666 AsnCysLeuMetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
 DB 2012 AACTGCTCATGTACAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGCCNAG 2071
 QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheLeu 705
 DB 2072 CGTCTCCAGGCTTTGACCGTGTCCGAAATGCTGAGATTGGAAATAAAGACTTTGAGCTT 2131
 QY 706 ThrTyrLeuGluGluAlaValTrpThrGluHisTrpLeuValArgIleValArgValLys 725
 DB 2132 GATGCTCTGGAGGAGGCTATACCAACAGACATTTGGCTGTGCTCAGATATACAGGTAAAG 2191
 RESULT 10
 ID AB232036 standard; DNA; 2256 BP.
 XX AB232036;
 XX AB232036;
 DT 30-JAN-2003 (first entry)
 XX
 DE Candida albicans essential gene SEQ ID NO 6323.
 KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 XX
 PD 11-JUL-2002.
 XX

PF 26-DEC-2001; 2001WO-US049486.
 XX
 PR 29-DEC-2000; 2000US-0259128P.
 PR 20-FEB-2001; 2001US-00732024.
 PR 22-AUG-2001; 2001US-0314050P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Roemer T, Jiang B, Boone C, Bussey H, Ohlson KL;
 DR WPI: 2002-566694/60.
 DR P-PSDB; ABP73486.
 XX
 PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele of
 PT a gene and placing other allele of the gene under conditional expression.
 XX
 PS Claim 37; SEQ ID NO 6323; 167pp + Sequence Listing; English.
 CC
 CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office
 XX
 SQ Sequence 2256 BP; 648 A; 336 C; 456 G; 816 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.35e-212 Length: 2256
 Score: 2175.00 Matches: 434
 Percent Similarity: 70.16% Conservative: 109
 Best Local Similarity: 56.07% Mismatches: 169
 Query Match: 53.78% Indels: 62
 DB: 6 Gaps: 15

US-10-028-384-8 (1-774) x AB232036 (1-2256)

QY 11 LysValAlaGlyTyrSer---SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
 DB 34 AAAGTATGGGATAGATGTTGAAGCTATTAGAGCTTTTATTAAGATTATATTATA 93
 QY 30 -----AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIle 45
 DB 94 TCGATAGCAGGTGCAGCTATTCTTCCTCGTTTATTTCCGTGATTCGATTGGAAGTATT 153
 QY 46 IleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaValMetValGlnAsnGly 65
 DB 154 ATTATGAATTCGATCCCTTGTTCAATTCGACCAACCAATATTATTTGACTCATTC 213
 QY 66 TrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleVal 85
 DB 214 TTTTATGAATTTTGAATTTGTTTATGATGATAGAACTTGGTACCCATTGGGAAGACT 273

QY 86 GlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisThrLeuHis 105
 DB 274 GGTGGTACTTATATCCGGTTTAAATGGTGACCTCAGGTGCCATT---TGG-----CAT 324
 QY 106 ValLeuAsn-----IleProValHisIleAtrAspIleCysValPheLeuAla 121
 DB 325 ATTTTCGTGATGGTTGGCTTACCCTGTGATATTAGAAATATTGGTTTATTAGCA 384
 QY 122 ProllePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrGlySerLeuPheSer 141
 DB 385 CCAGTTTCTCGGGATTAACGTCAATTTGTACTTATTTTTCGACATAAGCAATGAAGAT 444
 QY 142 AlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArg 161
 DB 445 TCTAGTCAGGATTTATGGCAGCTATATTATGGGATTTGCCAGGTTATATTTCAGA 504
 QY 162 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
 DB 505 TCAGTGGCTGGTCTTATGATATAATGAAGCAATTCGCAATTTATTTATGCGCAACATTT 564
 QY 182 PheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaLeu 201
 DB 565 TATTTCTGGATTAATCAATGAATATGGGTTCAGTTTCTATGCCACATTCAGCAITA 624
 QY 202 SerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleAlaLeuLeuPro 221
 DB 625 TTCTATTCTATATGTTAGTGGTGGTGGATATGTTTTCATTTACCAATTTGATTCGA 684
 QY 222 LeuHisValPheValLeuLeuIleMetGlyValGlyTyrSerProArgLeuLeuThrSerTyr 241
 DB 685 TTACAGCTATTGTCCTGATTTTCATGGGTCGGTTCAGTTTATGCGCAACTTTACACTGCTAT 744
 QY 242 SerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGln 261
 DB 745 ACTACATGTTGCTTGGGTACTTTGGCATCAATGCGATTCATTCCTGGGTTTTTA 804
 QY 262 ProlleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaVal 281
 DB 805 CCAATTAGATCAATGATCATATGGCTGCGCATTTAGGAGTATTTGGATTTGTACAAATTAGT 864
 QY 282 AlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPheIle 301
 DB 865 GCTTTTGGTGATTTGTTAAATCAAAAGTTCCAAACCAAAATTTAAATCATTTCTTGATA 924
 QY 302 ValGly---GlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 925 GTTTCATTTGCTGTGTGGATTTAGGTATTT---GGTGGATTTATTTGGATTAAACACA 981
 QY 321 LeuGlyValValAlaProTyrSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 982 ATGGGTTGGATGCTCTCGACAGGTAGATT---TATCTTATGGATACAAATTTATGCC 1041
 QY 341 LysIleHisProlleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 1042 AAGATTCAATTCATATATGCTTCTGTTTCGAAACATCAACCTACTGCTGGCCAGCA 1101
 QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
 DB 1102 TTTCTTTTCGATAGTATGCTTATTTGGTTATTTCCCGCTGGTATCTATTATGTTTC 1161
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1162 CAGAATTGAAGATGAACAGTTTTCATTTATCATTTACAGTATTTGTTGTTCTATTATT 1221
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1222 GCTGGTGTGATGTTAGATTCATGTTGACTTTGACTCCAGTCATTTGTTGCTCGACGA 1281
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
 DB 1282 ATTGCCCTTAATCAATTTGTTGATGCTATTG-----AGAAACAGATCCCAT 2226
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThr 460

DB 1315 -----GACATTTGTTGATTTGTTCACT 1335
 QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThr 480
 DB 1336 -----GAGAAGTTGGAAAGTATCATCATGACCTAGTCAGCAATCCAGAGCTCAACC 1389
 QY 481 GlyValSerSerAsn-----LeuLysSerIleValIleLeu 492
 DB 1390 AAAAAATCAAGTTCCAGATTTCCAAATGCTGATATTGTTCAAAAGTTTGGTTTACTG 1449
 QY 493 AlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAla 512
 DB 1450 ACATTTACATTTTACTTTTCTTACTTTTACTTTTACTTTGGTAAACATCGAATGCT 1509
 QY 513 TyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIle 532
 DB 1510 TATTCATCACCATCAGTTGTTTATGCAATCCAGAAC---CCAGATGGCTCACAACATATC 1566
 QY 533 LeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArg 552
 DB 1567 ATGTATGATATAGAGAAGCTTATCTGTTTAAAGATGAATACACAGAGATGCCAAA 1626
 QY 553 ValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 572
 DB 1627 GTTATGGCTGGTGGATTTATGTTTATCAATCGGGGTATGGCTGATAGAACACACATT 1686
 QY 573 ValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
 DB 1687 GTTGTATCAATACATGGAATAACACATATTCGCCTGTTGTAAGCAATGCTCTCC 1746
 QY 593 ThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
 DB 1747 CTTGAAGATGTCGTGTAATTTTGAACAACACCATGTTGATTATGTTAGTATTATA 1806
 QY 613 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArg 632
 DB 1807 TTGGNGGTTATGGGTATTTCTGGTATGATATTAAACAATTTTATGATGATGAAGA 1866
 QY 633 IleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGly 652
 DB 1867 ATGTCTGAGGTATCTGGCTGATGAATCAAGAAAGAGAGACTACTTTACTGCCGAGGA 1926
 QY 653 GluPheArgValAspAlaGluValAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeu 672
 DB 1927 GAATATAAGTGGATAAAGATGCATCCTGCAATGAAGATTTCTTTGATGATAGTTA 1986
 QY 673 SerTyrTrpArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArg 692
 DB 1987 TCGTATCATAGATTTCATGAATTTG-----TTTGGAGGTAGAGATGGTGTGATAGA 2037
 QY 693 ThrArgAsnAlaValIle---GlyAsnLysAspPheAspLeuThrTyrLeuGluGluAla 711
 DB 2038 GTTAGAACCAACAATCCAGCCCAATGAAGTACCGAAATGAATGTTGTTGAAGAGCC 2097
 QY 712 TyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHisGluPheAsn 731
 DB 2098 TTCATATCAGAAAAATTTGGATTTGAGAAATTTACAAAGTTAAA----- 2139
 QY 732 ArgProSerLeuLysThrLysGluArgThrIleProAlaAsnPheIleSerArgLys 751
 DB 2140 -----GATTTGGATAATGTTGGTAGAGATTACATCAAGCTACTGCTTTTGAAGATCA 2193
 QY 752 AsnSerLysArgArgGlyGlyTyrIleArgAsnArgProVal 765
 DB 2194 TCATCGGCACCTTCCAAA-----AGAAACAGATCCCAT 2226

RESULT 11
 ABT20789
 ID ABT20789 standard; DNA; 2232 BP.
 XX AC ABT20789;
 XX

DT 16-APR-2003 (first entry)
DE Aspergillus fumigatus essential gene #3147.
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
XX Aspergillus fumigatus.
OS WO200286090-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US013142.
XX 23-APR-2001; 2001US-0285697P.
XX 27-APR-2001; 2001US-0287066P.
XX 05-JUN-2001; 2001US-0295890P.
XX 09-JUL-2001; 2001US-0303899P.
XX 31-AUG-2001; 2001US-0316362P.
XX (ELIT-) ELITRA PHARM INC.
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX Disclosure; Page; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterization, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention
XX
SQ Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 7,52e-212 Length: 2232
Score: 2172.50 Matches: 427
Percent Similarity: 71.24% Conservative: 108
Best Local Similarity: 56.86% Mismatches: 171
Query Match: 53.70% Indels: 45
DB: 8 Gaps: 11
US-10-028-384-8 (1-774) x ABr20789 (1-2232)
QY 18 LeuLeuThrPheAlaIleLeuLeuLeuAlaTrpLeuAlaGlyPheSerArgLeuPhe 37

61 CTTCTGGGATTATATCTCTGTGTACCATTCGACGACGACGATTCGACGCTCTTC 120
38 AlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTrpArgAla 57
121 AGCGTTATCGGCTTCGAGAGTATCATCCAGCAAGTTCACCCGTGGTTCACATTCGAGCA 180
58 ThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArgAla 77
181 ACATAATACTAGTACAGATGGTTCTATAGCTTTGGGATGGTTGTATGATGAGGAACA 240
78 TrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThrSer 97
241 TGGCATCTCTCGGACGCTCACCCTGGGACGCTTATATCCCGTCTCATCGGTGACGAGC 300
98 GlyGlyIleHisTrpLeuHisValLeuAsnIleProValHisIleArgAspIleCys 117
301 GCGGTGATCTACCATATCTTGGGATCTCTACTATCCCTCGATATTCGACATCTGC 360
118 ValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLys 137
361 GTCTTACTGGCGCCAGGATTCTCCGGCTGACTGATTCGCAATGACTTGTGTCGATCC 420
138 GluLeuTrp---SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
421 GAGATGTCCTCTGGCATCTGCGAGCTCTTCTGCGACGACCTTCATGGGATCGCCCT 480
157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
481 GGTATACATCTCCGATCAGTTGCTGGAGCTAGCATAAACGAGCATGTGCACTTTCTG 540
177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
541 CTTGTTTTCATCTCTTCTATGGATCAGCTGTCAAAATGGGTCTATCATCTGGGGA 600
187 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
601 GCGTTCGACGACATCTCTACGGCTACATGGTGTGCGCATGGGTGGTATGTCTCAT 660
217 IleAsnLeuIleProLeuHisValPheValLeuLeuMetGlyArgTyrSerProArg 236
661 ACGAACCTGATCCCTCGACGTTTTTGTCTTCTGTCATGGGTAGATACAGCATCGC 720
237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
721 ATCTACATAGTATATACCATATGATGGTATGCGCTGGGACTTTGGCTAGCATCGAGTCC 780
257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
781 TTCGTGGATTTTGGCTATCCGAAACGACGACCATGTCGCCCTTGGGTGTCTTCGGC 840
277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
841 CTGCTTCAGCTTGTGGCTTCGCGAGTTTGTCCGACGCTTCGTTCCAAAGCAAGCAGTTC 900
297 ArgLysLeu-----PheIleValGlyGlyLeu---LeuValGlyValGly 310
901 CAGAGATCTTCGACCGCCATGATCTTCATCACTTCGCTCGTCTCGTTGTGCG----- 954
311 ValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArg 330
955 -----CTAGTTGTTCTGACTGTGCGGAGTGTATCGCTCTCTGGAGCGCGCA 1002
331 PheTyrSerLeuTrpAspThrGlyTyrAlaValIleHisIleProIleIleAlaSerVal 350
1003 TTCTACTCTTGTGGGACACTGGCTATGCCAAATCCAAATCCAAATTCCTCATCTGCTCAGTC 1062
351 SerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCys 370
1063 TCGGAACACACAGCCACCGCTTGGCCAGCGTCTCTCTTCGATCTGAATCTCCTGATCTGG 1122
371 AlaPheProValGlyValTrpTyrCysIleGlnIleAsnAspGluArgValPheVal 390

| | | | |
|------|----|---|------|
| 1123 | Db | CTTTTCCGGGAGGCTGTCTACATGTCCTTCCGTGACCTCAAGGACGAGCATGCTTCCTGCT | 1182 |
| 331 | Qy | ValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThr | 410 |
| 1183 | Db | ATTATTACTCGGTCTTCGGAGCTACTTTCGCGCGGTATTATGGTCCGACTAATGCTGACC | 1242 |
| 411 | Qy | LeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPhe | 430 |
| 1243 | Db | TTGACCCCTATTGTGTGTGTGCGCGTGTCTGGCGCTGTCTGTCATCTCGACACCTAT | 1302 |
| 431 | Qy | LeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAsp | 450 |
| 1303 | Db | ATGGCGACT-----ACCTCCCGACACCAACGGTCTGGAAGCGGAA | 1341 |
| 451 | Qy | GluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHis | 470 |
| 1342 | Db | ACGAATGAAGACTCGTTCACAACT----- | 1368 |
| 471 | Qy | ArgThrLysHisAspAlaGlnInAspThrGlyValSerSerAsnLeuLysSerIleVal | 490 |
| 1369 | Db | ---CTTCGCTCAGTTCGGAAGCCCAATGTTGGAATCACTCCCATGTTTCTAAGATTATA | 1425 |
| 491 | Qy | IleLeuAla---ValLeuMetLeuLeuMetMetPheAlaValHisCysThrTdpValThr | 509 |
| 1426 | Db | GTGACGGCGTCTGTTGTCTACTGCTCTCTGTTGTCGCACTGCACCTGGGTACA | 1485 |
| 510 | Qy | SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer | 529 |
| 1486 | Db | TCGAATGCATACTCTCTCCTTCGTTGCTTCTGGCT---AGTCGGATGCCTCGACGAAGC | 1542 |
| 530 | Qy | ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTdpLeuSerGlnAsnThrAlaAsp | 549 |
| 1543 | Db | CAATACATCATTCAGCATATCTGAGGCTTACTGCTTCTGTCAGATATCTCCTCAG | 1602 |
| 550 | Qy | AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg | 569 |
| 1603 | Db | AACGCCAAAATCATGTATGCTGGGATATGGGTATCAAAATCGGTGGCATGGCGGACCGC | 1662 |
| 570 | Qy | ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla | 589 |
| 1663 | Db | CCAACTTGGTTGACAACACACACTCGAACACACCAATATTGCTACGGTTGTTAGCGG | 1722 |
| 590 | Qy | MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal | 609 |
| 1723 | Db | ATGAGCTCAGCGGAGGATCACTACCCCATCTCCGCCAGCATGATGTCGATACGATG | 1782 |
| 610 | Qy | LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp | 629 |
| 1783 | Db | CTGGTGGTGTTCGGTGTCTGTAGGTATTCTGGGATGACATTACAAATCTTATGCG | 1842 |
| 630 | Qy | MetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThr | 649 |
| 1843 | Db | ATGTCCTGATATCGCCGAAGGTATCTGCCCGCATGAGGTAAAGACGGGACTCTCTTACT | 1902 |
| 650 | Qy | AspArgGlyGluPheArgValAspAlaGluGlyAlaIaPcAlaLeuLeuAsnCysLeuMet | 669 |
| 1903 | Db | GCACGCGTGAATATCTGTGACGATGGAGCGACCCCAACTATGTCGCAACAGCTGTGG | 1962 |
| 670 | Qy | TyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGly | 689 |
| 1963 | Db | TATAAATGTCTTATACAAATTTCAACTCTCTC-----TTCCGCTCGGGC | 2007 |
| 690 | Qy | -----TyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyr | 707 |
| 2008 | Db | CAAGTGTGCAACGCGCTCGTGGTCAAACCTCCACAGAAGGCGCCCTCAGCTCTCTACA | 2067 |
| 708 | Qy | LeuGluGluAlaTyrThrThrGluHisTdpLeuValArgIleTyrArgValLysPro | 727 |
| 2068 | Db | CTCAAGAAGCTTTCACGACCGGAGATCGATTCATTCGTATCTACAGGTCAGAGATCTT | 2127 |
| 728 | Qy | HisGluPheAsnArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPhe | 747 |
| 2128 | Db | GACAACTTGGCCGA-----GACCAACAACCGAGTGTGCTTCGACAAAGGT | 2175 |

```

748 lIeSerArgIysAsnSerIysArgIysGly 758
::: :::: ||||| ::::|
2176 CTCAGAAAAGCGGAGTACAAAGGAAAGGGA 2208

RESULT 12
ABV24502
ID ABV24502 standard; cDNA; 2760 BP.
XX AC
XX ABV24502;
XX DT
XX 16-SEP-2002 (first entry)
XX DE
XX Human prostate expression marker cDNA 24493.
XX KW Human; prostate cancer; cytostatic; carcinogen;
XX KW Pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160960-A2.
XX PD
XX 23-AUG-2001.
XX PF
XX 20-FEB-2001; 2001WO-US005171.
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0319007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX DR
XX Novel isolated nucleic acid molecule associated
XX PPT prostate cells and correlating with presence of
XX PPT for detecting presence of prostate cancer, stages
XX PS Claim 1; Page 4632-4633; 11750pp; English.
XX CC The invention relates to an isolated nucleic acid
XX CC a nucleotide sequence given in Tables 1-9 (ABV000
XX CC specification or its complement. (I) is useful
XX CC a patient is afflicted with prostate cancer; (b)
XX CC progression of prostate cancer in a patient; (c)
XX CC of a test compound to inhibit prostate cancer; (d)
XX CC the efficacy of a therapy for inhibiting prostate
XX CC (e) selecting a composition for inhibiting prostate
XX CC (f) assessing the prostate cell carcinogenic pos
XX CC determining whether prostate cancer has metastas
XX CC assessing the aggressiveness or indolence of pro
XX CC ; (I) is also useful as a pharmacodynamic or ph
XX SQ Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0

Alignment Scores:
Pred. No.: 2,65e-211
Score: 2168.50
Percent Similarity: 73.27%
Best Local Similarity: 57.76%
Query Match: 53.60%
DB: 5

US-10-028-384-8 (1-774) x ABV24502 (1-2760)

17 SerLeulleThrPheAlaIleLeuLeulleAlaTrpLeu
::: :::: ||||| ::::|
150 ACATTTTAGGCTTCATCTGTCGTAAGGCTGCTGTA
::: :::: ||||| ::::|

```

QY 37 pheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrAr 56
Db 210 TTTTCTCTCTGAGATTGAAAGTGTATCCATGAGTTTGATCCGCTTAAATATATCG 269
QY 56 gAlaThrAlaTyrMetValGlnAsnGlyTyrTrpTyrAsnPheLeuAsnTrpPheAspGluAr 76
Db 270 GACTACACAGGTTCCTGGCTGAGGAGGGTITTAATAANTCCATAACTGGTTTGATGACCG 329
QY 76 gAlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleTh 96
Db 330 AGCCCTGGTACCCCTTTGGACCAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCAC 389
QY 96 rSerGlyGlyIleHisTrpLeuHis-ValLeuAsnIleProValHisIleArgAspI 116
Db 390 CTCTGCTCAATCTACCATGATCCATTTTTCACATCACCATCGACATTCGGAATG 449
QY 116 leCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuT 136
Db 450 TCTGTGTGTCTGCGCCCT 509
QY 136 hrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaLaserPheIleAlaIleValP 156
Db 510 CCARAGAGCTCAAGGATCGAGGGCTGGCTTCTGCTGCTGCCATGATTCGTAGTTC 569
QY 156 roGlyTyrIleSerhrGserValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheA 176
Db 570 CTGGATATATCTCCGATCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 629
QY 176 laLeuGlnPheThrTyrPheLeuTrpValArgSerValTyrGlySerValPheTrpS 196
Db 630 GCATGCTACTACCTACTACATGATGATGATCAAGGAGGATTAAGACTGGTTCCATCTGTTGG 689
QY 196 erAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheI 216
Db 690 CAGCTAAGTGTGCCCT 749
QY 216 leIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProA 236
Db 750 TGATCAACTTAATCTCTCCACGCTCTCTGCTGCTGATGCTCACAGGCGCTTCTCTCAC 809
QY 236 rGLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleP 256
Db 810 GGATCTATGTGGCT 869
QY 256 roPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheV 276
Db 870 CCTTTGTGGTTCAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 929
QY 276 alLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluP 296
Db 930 GTCTCTGCCAGATCCATCCCTTTGTGGATTACCTGGCAGCAATGGAATCCCAACAAT 989
QY 296 heArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyV 311
Db 990 TTGAAGTCTTTTCGGAGCGCTCATCTCTCTGTTAGCTTGTCTCTCTCACCGTGGAA- 1048
QY 311 alPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgP 331
Db 1049 -----GCTCTCTCATGTGACA-----GGAAAAAATATCTCTCCCTGGACGGCGCTT 1094
QY 331 heTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValS 351
Db 1095 TCTACTCGCTGCTGATCCCTCTTATGCTAAGAACACATCCCATCATTTCTTCTGTGT 1154
QY 351 erGluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysA 371
Db 1155 CTGAGCATCAGCCCAACACCTGGTCTCTCATATCTTTTGACCTTCGACCTCTCTCTCTCA 1214
QY 371 laPheProValGlyValTyrTyrCysIleLysGlnIleAsnAspGluArgValPheValV 391
Db 1215 TGTTCAGTGTGGCTCTTACTCTTACTCTTAGCAACCTGTCTGATGCGCGGATTTTATCA 1274

RESULT 13

QY 391 alLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgIleMetLeuThrL 411
Db 1275 TCATGTATGGTGTGACCATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1334
QY 411 euThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheL 431
Db 1335 TGGCACCTGTATGTGATCTCTCTGTCATCTCTGTCATCTCTGTCATCTCTGTCATCTCT 1394
QY 431 euGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspG 451
Db 1395 TGAAGATCTGGACATAGTCGCTCA-----GACAAAGAGCAAG----- 1420
QY 451 lAlaGluAspSerIleGlyLysThrLeuTyrAspLysAlaGlyLysLeuLysHisA 471
Db 1421 -----GACAAAGAGCAAG----- 1435
QY 471 rGThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerI 489
Db 1436 -----AAGCAACAGGATTCACCTACCTATTAAGAATGAAGTGGCAAGTG 1481
QY 489 leValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValT 509
Db 1482 GGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1541
QY 509 hrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlyS 529
Db 1542 CAGTGTAGGCTACT 1598
QY 529 erArGAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaA 549
Db 1599 GTAGATCATATTTGATGATCTTCGAGAGCATATATTGCTGCTCTCTCTCTCTCTCTCT 1658
QY 549 spAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnA 569
Db 1659 AGGATGGCAAGGTCT 1718
QY 569 rGThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysA 589
Db 1719 GAACAATTTTAGTGGACATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAG 1778
QY 589 laMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrV 609
Db 1779 CAATGGCGTCCACAGAGAGAAAGCCATGAGATCATGAGGAGCTCGAGTCTCGATG 1838
QY 609 alLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 629
Db 1839 TGCTGGTCAATTTTGGAGGCTCTACTGGGTATTTCTCTGATGATATCAACAAGTTTCTTT 1898
QY 629 roMetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluS 645
Db 1899 GGATGGTCCGATTTGGAGGAGACAGATACAGGCAACAT-----ATCAGAGAGA 1949
QY 645 erAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuL 665
Db 1950 ATGACTATTATCTCAACTGGGAGTTCCGTGGACCGTGAAGTTCTCTCAGTCTGC 2009
QY 665 euAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrA 685
Db 2010 TCAACTGCTCATGTACAGATGTTTACTATCGCTTGGCAGGTTTACACAGAGCCCA 2069
QY 685 rGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspL 705
Db 2070 AGCGTCTCCAGGCTTTGCGCGTGTCCGAAATGTGAGATTGGGAATAAAGACTTTGAGC 2129
QY 705 euThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValL 725
Db 2130 TTGATGCTCTGGAGAGCATATACACAGAACATTTGGCTGGTGGATATACAGGTAA 2189
QY 725 ys 725
Db 2190 AG 2191

ADB69900
 ID ADB69900 standard; DNA; 2466 BP.
 AC ADB69900;
 DT 04-DEC-2003 (first entry)
 DE C. neoformans open reading frame SEQ ID NO:2305.
 XX ds; gene; fungicide; gene therapy; infection.
 KW
 XX
 OS Cryptococcus neoformans.
 XX
 PN WO2003052076-A2.
 XX
 PD 26-JUN-2003.
 XX
 PF 17-DEC-2002; 2002MO-US040225.
 XX
 PR 17-DEC-2001; 2001US-0341251P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Zamudio C, Eroshkin AM;
 XX
 DR WPI; 2003-533017/50.
 XX
 DR P-FSDB; ADB70261.
 XX
 PT New nucleic acid, useful for preparing a composition for treating an
 XX infection caused by Cryptococcus neoformans.
 PS Claim 2; SEQ ID NO 2305; 136pp; English.
 XX
 CC The invention relates to a novel purified or isolated Cryptococcus
 CC neoformans nucleic acid molecule comprising a sequence encoding a
 CC polypeptide comprising a sequence not given in the specification. A
 CC polynucleotide of the invention has fungicide activity, and may have a
 CC use in gene therapy. The nucleic acid is useful for preparing a
 CC composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 9,678-206 Length: 2466
 Score: 2113.50 Matches: 407
 Percent Similarity: 70.82% Conservative: 122
 Best Local Similarity: 54.48% Mismatches: 175
 Query Match: 12 Indels: 43
 DB: 10 Gaps: 10
 US-10-028-384-8 (1-774) x ADB69900 (1-2466)
 QY 4 ThrProlyMetLeuAnSerIysValAlaGlyTyrSerSerLeuLeuThrPheAla 23
 DB 199 TCTCCGTCGACTATCAACACACG-----GAAAGCTTGTGGCTTCATCATC 246
 QY 24 LeuLeuLeuAlaThrLeuAlaGlyPheSerArgLeuPheAlaValIleArgPheGlu 43
 DB 247 CTGCGGTTAATATGCGGTGCTGGATGGAAGTCGATTGTTGGCGTATCAGATTGAA 306
 QY 44 SerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyrMetValGln 63
 DB 307 TCTGTCAATCCACGAATTTGACCCCTGGTTCACACTACCGAGCCTCGAAAGTTCTTGTAAAC 366
 QY 64 AsnGlyTrpTyrAsnPheLeuAnThrPheAspGluArgAlaTrpTyrProLeuGlyArg 83
 DB 367 AAGGGTTTCTACGAGTTCTGGAACACTGGTTTGACCCCTCGCTTGTGACCTCTCGGAGA 426
 QY 84 IleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyIleHisTrpLeu 103

DB 427 ACTGCGGTACACCGCTCTATCTTGGCTTGATGTGTACGCTCGTGCATGATTTGGCATGCT 486
 QY 104 LeuHisValLeuAnSerIleProValHisIleArgAspIleCysValPheLeuAlaProIle 123
 DB 487 CTTCCGGCAATCAATATGCCCGTGGACATTCGCAATCTGTGTCTCTCTCCCTTGACCTGGA 546
 QY 124 PheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrIysGluLeuLeuTrpSerAlaGly 143
 DB 547 TTTTCTGGATTGACTGCTGGCGACTTATCTTTTCCCACTGAATGTCTACACCATCA 606
 QY 144 AlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArgSerVal 163
 DB 607 GCTGGTCTATTGGCGCGCGCTTTCATTGGCATTTACCGGATACATCTCTCGATCTGTC 666
 QY 164 AlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeu 183
 DB 667 GCGGTCTCTATGACACGAAGCCATTCCTCTGTGTGATGAGCTCCTTCTACTCT 726
 QY 184 TrpValArgSerValIysThrGlySerValPheThrSerAlaAlaAlaLeuSerTyr 203
 DB 727 TGGATTAAAGCGCGTCAAAACCGGTAGCTCAATTTGGGTATGATCACTGCTTGTCTAC 786
 QY 204 PheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIleProLeuHis 223
 DB 787 GGGTGGATGGTGTGTCATGGGTGGTGTGATCTTTCATCAACACATGATTCATTCGAC 846
 QY 224 ValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSerTyrThr 243
 DB 847 GCCTTTGTCTCATGTGTCATGGCAGGTTCACCAACCGCTTTATACCGCTTACTCTTCC 906
 QY 244 PheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGlnProIle 263
 DB 907 TGGTATGTCATTGGAACTATCGCTTCATGAGTCCCTTTTGGAGTTCTCCCCCATC 966
 QY 264 ArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThr 283
 DB 967 CGAACCTCTGAGCATGCGCGCTTGGGTGTGTTTGGTCTGTGTACAGCTGATCGGATTC 1026
 QY 284 LeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArg-----LysLeuPhe 300
 DB 1027 GTCCGAAGTCGTCCGACGATCTGCTCCGCGCAAGCAATTCAGCTCCTTCACAAAGTTT 1086
 QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 DB 1087 CTCGTGGCGGATTCCTGCTCAGT-----TTTGTGCGCTCGTCACTTTGACTTTC 1137
 QY 321 LeuGlyValValAlaProThrPheSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
 DB 1138 TCTGGATGGATGCGCCCTTCGCTGGAAGATTTTATTCTCTTTGGGATAGTGGCTATGG 1197
 QY 341 LysIleHisIleProIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
 DB 1198 AAGGTCCACATGCCCATTTATGCTCGCTCTCGAACACACAGCCCGCTTGGCGCTCA 1257
 QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrPheCysIle 380
 DB 1258 TTTCTACTTGTACCTCGAAATGCTTATCTTTTCCCTCCCGGTGCTTCTGGTGTTC 1317
 QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
 DB 1318 AAGGAGCTTCGCGATGAGCAGATCTTCATCATCATTTATGCGGTTCGCTGCTATTTT 1377
 QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 DB 1378 GCGGTGTCTGTTGCGATGCTTATGCTGTGTCATCAGCTTGTGTGTGTCTCTCGGCC 1437
 QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerIysArgMetGly 440
 DB 1438 ATTGGGTTCTCCAAACTTCTCGAGCGGTATATTGAC----- 1473
 QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGlyIleThr 460

Db 372 TATAGGCTACCAATATCTCGTCAACAAATTCGTTTACAAAGTTTTTGAAGCTGGTTGAC 431
Qy 75 GluArgAlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMet 94
Db 432 GACCGTACCTGGTACCCCTCGAGAGGTACTGGAGGACCTTATATCTCGTTGATG 491
Qy 95 IleThrSerGlyGlyle---HisTrpLeuLeuHisValLeuAsnIleProValHisIle 113
Db 492 ACGACTAGTCGTATCTGGCAGCCCTCGCAACTGGTGGCTTCCCATTCACATC 551
Qy 114 ArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyr 133
Db 552 AGAACCCTTGTGCTATTGTGGCCACTATTCTGGGGTCCCGCTGGCGACTAC 611
Qy 134 LeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaIleSerPheIleAla 153
Db 612 GAATTTACGAAAGAGATTAAAGATGCCAGCGCTGGCGCTTTGGCTGCTTTATAGCC 671
Qy 154 IleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAla 173
Db 672 ATTGCCCCGGTTATATATCTAGATCAGTGGCGGGTCTTAGTAATAGAGCCATTGCC 731
Qy 174 IlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerVal 193
Db 732 ATTACACTATTAATGTGCTACTTTTCATGTTTGGATTAGGCCCAAAAGACTGGCTCTATC 791
Qy 194 PheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyr 213
Db 792 ATGCACGCAACGTGTCAGCTTATCTACTCTTACATGGTGTGGCTGGGTGGATAC 851
Qy 214 ValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyr 233
Db 852 GTGTTCATCAACCACTGATCCCATCTCCATGCTTTTGGCTGATTGATGGGAGATAT 911
Qy 234 SerProArgLeuLeuSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMet 253
Db 912 TCGTCCAACTGATTCTGCTACACCACTTGGTAGCTTGGTGGTGGTGGTGGTGGT 971
Qy 254 GlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGly 273
Db 972 CAGATCCCATTTGCGGTTTCTTACCTATCAGTCTTAACGACCATCGCCGCTATGGGT 1031
Qy 274 ValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArg 293
Db 1032 GTTTTCGGTTGATTGATTCAGATTGTCGCTTCGGTGGTGGTGGTGGTGGTGGT 1091
Qy 294 AsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyValPheVal 313
Db 1092 GCTAAGTTTAAAGTCATCATGATGTTCTCTGTTTGTGTTGATCTGCTGCTGCTGCT 1151
Qy 314 AlaValValLeuLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPheTyrSer 333
Db 1152 GGACTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1211
Qy 334 LeuTrpAspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGluHis 353
Db 1212 TTATGGGATACCAACTACCAAGATCCACATCTCTATCATTCCTGCTGCTGCTGCTGCT 1271
Qy 354 GlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAlaPhePro 373
Db 1272 CAACCGGTTTCGGTGGCGGCTTCTTCTTTGATACCACTTTTGTGCTGCTATTCCTC 1331
Qy 374 ValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValValLeuTyr 393
Db 1332 CGCGGTGTATCTCTACTATTCCTCGACTTGAAGACGACGCTTTTGTCTGCTGCTTAC 1391
Qy 394 AlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrPro 413
Db 1392 TCGGTTCTGTTGTTGCTATTGTCGGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1451
Qy 414 ValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGlu 433
Db 1452 GTCATCTGTGTGTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1511

RESULT 15

ABT20191

ID ABT20191 standard, DNA, 2603 BP.

XX

Qy 434 AspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAlaGlu 453
Db 1512 AAGACAAGTGACCGC----- 1526
Qy 454 AspSerIleGluLysLeuThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThrLys 473
Db 1527 -----AAATACGCCATCAAA 1541
Qy 474 HisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAla 493
Db 1542 CCTGGCGGACTACTG-----GCCAAATGATTGTTCCGGATCA 1580
Qy 494 ValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThrSerAsnAlaTyr 513
Db 1581 TTCATCTTTTATTTGATCTCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1640
Qy 514 SerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeu 533
Db 1641 TCTTCT 1697
Qy 534 AspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgVal 553
Db 1698 GACGACTTCAGGGAAGCGTACTATTGGTTAAGAATGAACCTCTGATGAGGACAGTAAGGT 1757
Qy 554 MetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuVal 573
Db 1758 GCAGCGTGTGGGATTACGGTTTCCAAATGGTGGCATGCGACAGACACCACTTTAGTC 1817
Qy 574 AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerThr 593
Db 1818 GATACACACACGTGGAACAATACATCAGATCGCCATCGTGTGTAAGCCATGGCTTCCCT 1877
Qy 594 GluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePhe 613
Db 1878 GAAGAGAAATCTTACGAAATCTTAAAGAGCATGATCGATTATGTTCTTGTGTCATCTTT 1937
Qy 614 GlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIle 633
Db 1938 GTGTGCTAAATGGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1997
Qy 634 AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGlyGlu 653
Db 1998 ACCGAGGAATCTGCCAGAGAGATAAAGAGCGTTATTTCATACCGCAGAGGAGAA 2057
Qy 654 PheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSer 673
Db 2058 TACAGAGTAGATGCAAGGGCTTCTGAGACCATGAGGAACCTCGTACTTTTACAAGATGTC 2117
Qy 674 TyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspA-gthr 693
Db 2118 TACAAAGATTTCCACAAATTA-----TTCAATGTCGCCCAAGCCACTGACAGATG 2168
Qy 694 ArgAsnAlaValIleGlyAsnLysAspPhe---AspLeuThrTyrLeuGluAlaTyr 712
Db 2169 CGTCAACAAATGATCACACCATTAGACGTCGCCCATTAGACTACTTCGACGAAAGTTTTT 2228
Qy 713 ThrThrGluHisTrpLeuValArgIleTyrArgValLysLysProHisGluPheAsnArg 732
Db 2229 ACTTCGGAATACTGGATGTTAGATATATCAATTTGAAGAGGAT----- 2273
Qy 733 ProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIleSerArgLysAsn 752
Db 2274 -----GATGCCCAAGGTAGAACTTTGAGGAGCGTTGGTGGATTAACCGGTTCTCT 2324
Qy 753 SerLysArgArgLysGlyTyrIleArg 761
Db 2325 ACGAAACACGAGGTCCTCCATAAAGAGA 2351

| | | | | | |
|------|----|--|---|--|------|
| 284 | QY | euA | gHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeu | ----- | 2299 |
| 1170 | Db | CCGAGTTGTGTC | CGGAGCTTGGTTC | CCAAAGCAAGCAGTTCGACGACGCTTCGACCGCCATGA | 1229 |
| 300 | QY | --PheIleValGlyGlyLeu | --LeuValGlyValGlyValPheValAlaValValVal | 318 | |
| 1230 | Db | TCATCATCACCT | CGGTCTCGGTTGGT | -----CTAGTTGTTCT | 1271 |
| 318 | QY | euThrMetLeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrG | 338 | | |
| 1272 | Db | TGACTGTGACGGGAGTATCGCTCCCTTGAGCGGGCGGATCTTACTCTCTGTGGACACTG | 1331 | | |
| 338 | QY | lyTyAlaLysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThr | 358 | | |
| 1332 | Db | GCTATGCCAAATCCACATTCCTCCATTCATTCCTCAGCTCGGACACCGCCACCGCTT | 1391 | | |
| 358 | QY | rpPheSerPhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpT | 378 | | |
| 1392 | Db | GGCCAGCGGTCTCTTCGATCTGAATCTCTGATCTGCTTTTCCGGCAGGTGCTTACA | 1451 | | |
| 378 | QY | yrCysIleLysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlav | 398 | | |
| 1452 | Db | TGTGCTCCGFGACCTCAAGGACGACATGTCTTGTCATTTACTTACTCGTCCCTGGCA | 1511 | | |
| 398 | QY | alTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetL | 418 | | |
| 1512 | Db | GCTACTTCCGCGGTGTATGGTCCGACTAATGCTGACCTTGACCCCTATTGTGTGTGTG | 1571 | | |
| 418 | QY | euAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysA | 438 | | |
| 1572 | Db | CGCTGCTCTGGCGCTGTGCTCCATCTCGACACCTATATGCGCACT | 1618 | | |
| 438 | QY | rgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluL | 458 | | |
| 1619 | Db | -----ACCCTCCGACGAGCGCTCGAAGCAACGAATGAGACTCGTCTTCAC | 1670 | | |
| 458 | QY | ysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnG | 478 | | |
| 1671 | Db | AACTCT | -----CTTCGCTCAGTTCGGAAGC | 1694 | |
| 478 | QY | lnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAla | 497 | | |
| 1695 | Db | CCATGTGGATCACCCTCCATGTCTTCAAGATTATGATGACGGCGTCTGTGCTGCT | 1754 | | |
| 497 | QY | euLeuMetMetPheAlaValHisCysThrTrpValTyrSerAsnAlaTyrSerSerPro | 517 | | |
| 1755 | Db | ACCTGCTCCTGTGTGTGCGCACTGCACCTCGGTGTACATCGAATGCATCTCTCTCCTT | 1814 | | |
| 517 | QY | erIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheA | 537 | | |
| 1815 | Db | CGGTGTCTGGCT | ---AGTCGATGCTCGCGGAAGCCATCATCATATGACGATTATC | 1871 | |
| 537 | QY | rgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArgValMetSerTrpT | 557 | | |
| 1872 | Db | GTGAGGCTTACTACTGCTTCTGTCAGAACTCCTCAGAACGCCAAATCATGTGTCATGGT | 1931 | | |
| 557 | QY | rpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspAsnAsnt | 577 | | |
| 1932 | Db | GGGATTATGGGTATCAATCGGTGGCATCGCGGACCGCCCAACTTGGTTGACACACACA | 1991 | | |
| 577 | QY | hrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerSerThrGluGluLys | 597 | | |
| 1992 | Db | CCTGGAAACAACCCATATTGCTACCGTTGGTAAGCGCATGAGCTCACGGCGAAGATCA | 2051 | | |
| 597 | QY | erTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePheGlyValIle | 617 | | |
| 2052 | Db | GCTACCCCATCTCCGCCACATGATTCGATTACGTGCTGGTGGTGTTCGGTGGTCTGC | 2111 | | |
| 617 | QY | leGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyG | 637 | | |
| 2112 | Db | TAGGTATTCTCTGGCGATGACATTAACAAAATCTTATGATGTCGTATGCGCAAGGTA | 2171 | | |
| 637 | QY | luHisProLysAspIleLysGluSerAspTyrPheThrAspGlyGlyLysPheArgValA | 657 | | |

Search completed: December 14, 2004, 20:05:38
Job time : 1020.03 secs

Db 2107 ACGAAACACGAGGTCCTAAGAGA 2133

RESULT 2
US-09-270-767-12331
; Sequence 12331, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-034
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12331
; LENGTH: 900
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-12331

Alignment Scores:
Pred. No.: 1,666-160 Length: 900
Score: 1435.00 Matches: 273
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 35.47% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-12331 (1-900)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 82 ATGATCGGACGCCAAGATGCTGAACAGCAAGGTGGCTGCTACAGCAGCTAATCAC 141
QY 21 PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerArgLeuPheAlaValIle 40
Db 142 TTGGCATCTCTGTAATGCGTGGCTGGCGGATTTCTCTCGCCCTCTCGCGGTATC 201
QY 41 ArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyr 60
Db 202 CGTTTCAGTGCATATCCATGAGTTTGATCCGTGGTTCACTACCGGGCCACCGCTAC 261
QY 61 MetValGlnAsnGlyTrpTyrAsnPheIleuAsnTrpPheAspGluArgAlaTrpTyrPro 80
Db 262 ATGGTGCAGATGGTGGTGGTACAACTTCTCACTGGTTTCGACGAGCGCGATGATCCG 321
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
Db 322 CTCGGCAGGATGTGGCGGTACCGCTATCCCGCGCTGATGATTACGTCCGGCGGAATC 381
QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 382 CATTGGCTGCTCAGCTACTCAACATACCGGTGCATATCGTGACATCGCGTTCCTG 441
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
Db 442 GCGCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACCAAGAGCTGTGG 501
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 502 TCGCGGGCGCGGCTCTTCGCGCGCAGCTTCATCGCCATCTGGCTGGCTACATCAGT 561
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 562 AGGTGCGTGGCTGGATCGTACGATAACAGAGGCGATTGGCATATTCGCGCTCGATTCAC 621
QY 181 TyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSerAlaAlaAla 200
Db 622 TACTTCCTGTGGTGGCTGAGTGAAGTGAATCGGTCTGCTGGCGCGAGCGCT 681
QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIleIleAsnLeuIle 220
Db 682 TTGCTCTACTTCTACATGTTGCGCCCTGGGGTGGCTACGTTGTTCAATCAATCAATGATA 741

QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
Db 742 CCCCTGCACGCTTCGTACTGCTCATTTATGGGAGGTACTCGCGCGGTCTGCTGACACG 801
QY 241 TyrSerThrPheTyrIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPhe 260
Db 802 TACAGCACCTTCTACATCTCTGGGACCTGCTGTTCTCCATGAGATCCCTTCGTGGGATC 861
QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGly 273
Db 862 CAACCGATACGACACAGTGAACACATGCTGCGCTGGGA 900

RESULT 3
US-09-270-767-158
; Sequence 158, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-158

Alignment Scores:
Pred. No.: 4,48-147 Length: 1660
Score: 1325.00 Matches: 305
Percent Similarity: 62.48% Conservative: 38
Best Local Similarity: 55.56% Mismatches: 121
Query Match: 32.77% Indels: 86
DB: 4 Gaps: 11

US-10-028-384-8 (1-774) x US-09-270-767-158 (1-1660)

QY 245 TyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGlnProIleArg 264
Db 80 TATGGCTATGTGTAAGTGTTTAAACAAAGTTATATATATATGTAAGATCAATACAA 139
QY 265 ThrSerGluHisMetAlaAlaLeuGlyValPhe 275
Db 140 TTA-----TTTTTTTTTGGGTTTATTTCAAGGTTTCGTTTCCCTTAAACAACAAC 190
QY 276 -----ValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArg 293
Db 191 CGTGGGTTTCGTATGTAGCCCTTTCGACGCTTCGA-----GTTCTTCTCGA 238
QY 294 AsnGlu-PheArg-----LysLeuPheIleValGlyLeuLeuValGln 308
Db 239 GATGAAGTTTCTGCTGGAGGAATCGTTCTCTCTGCTTTCAGTGATGGTCTATTGAATC 298
QY 308 yValGlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSe 328
Db 299 ATGGCGCTCTTTCACATATAGATGCGAACACCGAGTGTTCGTGNGTGA----- 349
QY 328 xGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleAl 348
Db 350 -----GGCCTCTCCAGGTAGTGCAGATCAAAAGTCTTATTTCCCAATCACCG 396
QY 348 aSerValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLe 368
Db 397 C-----GT 399
QY 368 uValCysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgVa 388
Db 400 TACGTGTCCGATC---ATATCCAGATGGGCTCTGTAGTCCAACTTCAATTCCTCCCAATC 456

QY 388 lPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMe 408
 DB 457 TGTAAGTACCTAATTTGTACATAGGAGCTTGAGCAGGCGCGACACCTTCGGCATCTA 516
 QY 408 tLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAs 428
 DB 517 CCCTGAATTCACCGCGCTGGTAAGTAATCGCTTTCCTTAATGCTTCCTGGATGCTCTC 576
 QY 428 pValPheLeuGluGlnAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGl 448
 DB 577 CCTC----- 580
 QY 448 uValAspGluAlaGluAspSerIleGlnLysThrLeuTyrAspLysAlaGlyLysLe 469
 DB 581 -----GGCAATTCGACCATCCACAGCACTTGTGTATCATCGCGCAAT 627
 QY 468 uLysHisArgThrLysHisAspAlaGlnAspThrGlyValSerAsnLeuLysSe 488
 DB 628 AGCGATCACACCGCCCAAGATCACCA-----AACGTAGTCCA 665
 QY 488 rIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpVa 508
 DB 667 CGTCAAGAGATGTCAATAATTCGTAGGACTTCTCCTCGGTTGAGACATTCG-----CT 720
 QY 508 lThrSerAsnAlaTyrSerSer-----ProSerIleValLeuAlaPheHisAsnSerGl 526
 DB 721 TGCCAACCGCGCTATGTGACTATTGTTCCAGTATTATATCCACTAGCGTCTGTCTGT 780
 QY 526 nAspGlySerArgAsnIleLeuAspPheArgGluAlaTyrTyrTrpLeuSerGlnAs 546
 DB 781 TTGCATTCCTCCGCAACATTTTACAGCATTTTCAGAGGCTTACTACTCGCTTTCGCGAA 840
 QY 546 nThrAlaAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMe 566
 DB 841 CACTGCCGATGCTCGCTGCTATGCTTGGTGGATACGATACCGATAGCGGAT 900
 QY 566 tAlaAsnArgThrThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuVa 586
 DB 901 GGCACACAGAACGACGCTAGTGGATTAATACGTGGAAACAATAGTACATAGCGCTGT 960
 QY 586 lGlyLysAlaMetSerThrGluGlnLysSerTyrGluIleMetThrSerLeuAspVa 606
 DB 961 TGCGAAGGCAATGCTTCACCGCAGGAGAGTCTTACGAAATATGACATCTCTGAGT 1020
 QY 606 lAspTyrValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLy 626
 DB 1021 GGACTAGCTTTGGTGTATCTTGGCGGTGTGATCGGCTATTCTCGCGATGATATACAA 1080
 QY 626 sPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAs 646
 DB 1081 GTTCTCTGGATGGTCCGATTCCTCAGGAGAGCATCCACAGGACATAGGAAGCGA 1140
 QY 646 pTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAs 666
 DB 1141 TTACTTTACCGCGCGGTGAATTCAGGGTAGATCCGAGGTGCTCGCGCCCTGCTCAA 1200
 QY 666 nCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGl 686
 DB 1201 CTGCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTTGAAGTTGGACTACAGAG 1260
 QY 686 yProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuTh 706
 DB 1261 TCCATCTGGATATGTCGACAGTAAACCGCGTCATTTGGGAATTAAGGACTTCGATCTGAC 1320
 QY 706 rTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLy 726
 DB 1321 CTACTCTGAGGAGCGCTTACACACAGACACTGGCTTGTTCGATCTATAGGTTGAAGA 1380
 QY 726 sProHisGluPheAsnArgProSerLeuLysThrLysGluArgThrIleProProAlaAs 746
 DB 1381 GCCCATGAGTTCATAGACATCTCTGAGACCAAGAGAGAAAGATTCCTCCAGCAAA 1440
 QY 746 nPheIleSerArgLysAsnSerLysArgLysGlyTyrIleArgAsnArgProValVa 766

DB 1441 CTTCAATTCGAGAAAGAACTCTAAGCTCGCAAGGGCTACATAGAAACCGACCGTGT 1500
 QY 766 lVallysglyLysArgThrLeuLys 774
 DB 1501 TGTTAAGGAAACGAACTTTGAAA 1525
 RESULT 4
 US-09-270-767-15440
 : Sequence 15440, Application US/09270767
 : Patent No. 6703491
 : GENERAL INFORMATION:
 : APPLICANT: Homburger et al.
 : TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 : FILE REFERENCE: File Reference: 7326-094
 : CURRENT APPLICATION NUMBER: US/09/270,767
 : CURRENT FILING DATE: 1999-03-17
 : NUMBER OF SEQ ID NOS: 62517
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 15440
 : LENGTH: 1660
 : TYPE: DNA
 : ORGANISM: *Drosophila melanogaster*
 : US-09-270-767-15440
 Alignment Scores:
 Pred. No.: 4, 4e-147 Length: 1660
 Score: 1326.00 Matches: 305
 Percent Similarity: 62.48% Conservative: 38
 Best Local Similarity: 55.56% Mismatches: 121
 Query Match: 32.77% Indels: 86
 Gaps: 11
 DB:
 US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)
 QY 245 TyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGlnProIleArg 264
 DB 80 TATGTGCTATGTAAAGTGTTTTAAACCAAGTATATATATATGTTGGTGAAGATCAATACAA 139
 QY 265 ThrSerGluHisMetAlaAlaLeuGlyValPhe----- 275
 DB 140 TTA-----TTTTTTTTTGGGTTTATTTCAAGGTTTCGTTTCCCTTAAACAACAAC 190
 QY 276 -----ValLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArg 293
 DB 191 CGGTGGGTTTCGTATGTAGCCCTTGCAGCGCTTCA-----GTCTTTCTCGA 238
 QY 294 AsnGlu-PheArg-----LysLeuPheIleValGlyLeuLeuValGl 308
 DB 239 GATGAAGTTTGTGGAGGAATCGTTCTCTCTGCTTTCAGTGTGATGCTATTGAATC 298
 QY 308 yValGlyValPheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSe 328
 DB 299 ATCGGCTTCTTCACATATAGATGCGAACAGCAGTGTCTGTGTGTA----- 349
 QY 328 rGlyArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleAl 348
 DB 350 -----GGCCTCTCTCCAGGTAGTGCAGATCAAAGTCTTATTCCCAATCACCG 396
 QY 348 aserValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLe 368
 DB 397 C-----GT 399
 QY 368 uValCysAlaPheProValGlyValTyrTyrCysIleLysGlnIleAsnAspGluArgVa 388
 DB 400 TACGTGTGCGATC---ATATCCAGATGGGCTCTGTAGTCCAACTCAATTCGCCGAATC 456
 QY 388 lPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMe 408
 DB 457 TGTAGTACCTTAATTTGTACATAGGCGATTGACAGGCGCGGAGCACCTTCGCGATCTA 516
 QY 408 tLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAs 428

QY 475 pAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAlaVal 495
 Db 1037 TCACCAAAAGTAGTCCACGTCAGAGATGTCATATTTTCGTAGGACTTCTCTCGG-TT 979
 QY 495 uMetLeuLeuMetMetPheAlaValHisCysThrTTPValThrSerAsnAlaTyrSerSe 515
 Db 978 GAAGACATTGCCCTGCCAACCGCTATGTGACTATTGTTCCAGTATTATTATCCACT 919
 QY 515 rProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAs 535
 Db 918 AGCGTCGTTCTGTTTGCATTCCTGCTATCT-----GGTATCCGTAATCCACCAAGAC 865
 QY 535 pPheArgGluAlaTyrTyrTyrTyrLeuSer-GlnAsnThrAlaAspAlaArgValMetS 555
 Db 864 ATACCGCGGATCATCGGAGTGTCTCGAAGCCA-----GTAGTAA 820
 QY 555 eTTPThrPaspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeuValAspA 575
 Db 819 GCCTCTCTGAATCGTCTAAATGTTGCGGAATGGCAAAACAGACGACGCTAGTGGATA 760
 QY 575 snAsnThrTTPAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluG 595
 Db 759 ATATACGTGGACATAGTACATAGCGCTGGTGGCAAGCAATGCTTCAACCGAGG 700
 QY 595 LuLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePheGlyG 615
 Db 699 AGAAGTCTCTACGAAATATGACATCTCTTGACGTGACTACGTTTGTGGTATCTTTGGCG 640
 QY 615 lYValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArgIleAlaG 635
 Db 639 GTGTGATCGCTATTCCGGCGATGATATACACAGTTCCTGTGGATGTCGCAATTTGCGG 580
 QY 635 LuGlyGluHisProLysAspIleGlySerAspTyrPheThrAspArgGlyGluPheA 655
 Db 579 AGGAGAGACATCCCAAGGACATTAAGGAAAGCGATTAATTTACCACCGCGGTGAATCA 520
 QY 655 rGValAspAlaGluGlyValAlaProAlaLeuAsnCysLeuMetTyrLysLeuSerTyrT 675
 Db 519 GGGTAGATCGGAAGGTGCTCCGCCCTGCTCAACTGCTTATGTACAAATTAAGTACT 460
 QY 675 YrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgA 695
 Db 459 ACAGATTCCGGGAATTGAAGTTGGACTACAGAGCCCACTCTGGATATATCGCACAGTA 400
 QY 695 snAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrArg 715
 Db 399 ACGCGGTCAITGGGAATAAGGACTTTTGATCTGACCTACCTGAGGAGGCTACACACAG 340
 QY 715 LuHisTTPLeuValArgIleTyrArgValLysLysProHisGluPheAsnArgProSerL 735
 Db 339 AACACTGGCTGTTCGCATCTATAGTGTGAAGAGCCGCATAGTTCATAGACCATCAC 280
 QY 735 euLysThrLysGluArgThrIleProAlaAsnPheIleSerArgLysAsnSerLysA 755
 Db 279 TGAAGACCAAGGAGAGACGATTCCTCCAGCAAACTTCATCTCGAAGAAAGAACTCGAAGC 220
 QY 755 rGArgLysGlyTyrIleArgAsnArgProValValLysGlyLysArgThrIleLys 774
 Db 219 GTCGCAAGGGCTACATACGAAACCGACCGGTTGTTTAAAGGAAACGACCTTGAA 161

RESULT 6

US-09-270-767-15440/c
 ; Sequence 15440, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15440

LENGTH: 1660
 TYPE: DNA
 ORGANISM: *Drosophila melanogaster*
 US-09-270-767-15440
 Alignment Scores:
 Prod. No.: 1-24e-121 Length: 1660
 Score: 1112.00 Matches: 274
 Percent Similarity: 58.73% Conservative: 32
 Best Local Similarity: 52.53% Mismatches: 136
 Query Match: 27.48% Indels: 81
 DB: 4 Gaps: 9
 US-10-028-384-8 (1-774) x US-09-270-767-15440 (1-1660)
 QY 272 LeuGlyValPhe-----ValLeuLeuMetAla 280
 Db 1534 TTGGGTTTATTTCAGAGTCTCTTTCCCTTAACAAACCGGTCGGTTTCGTATGTAGCC 1475
 QY 281 ValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGlu-PheArg----- 297
 Db 1474 CTTGGGAGCTTAGA-----GTTCTTTCTCAATGAAGTTTCTCGAGGAAT 1427
 QY 298 -----LysLeuPheIleValGlyLeuValGlyValGlyValPheValAlaVala 315
 Db 1426 CTTTCTCTCTTGGTCTTCAGTGTGCTTATTGAACCTCATGCGGCTTCTACCTATA 1367
 QY 315 lValValLeuThrMetLeuGlyValAlaProTyrSerGlyArgPheTyrSerLeuTr 335
 Db 1366 GATGCGCAACAGCCAGCTTCTGTGGTGA-----GGCCTCTCT 1329
 QY 335 pAspThrGlyTyrAlaLysIleHisIlePheIleAlaSerValSerGluHisGlnPr 355
 Db 1328 CCAGTAGTCAGATCGAAGTCCTTATCCCAATGACGCG----- 1289
 QY 355 oThrThrTyrPheSerPhePheAspLeuHisIleLeuValCysAlaPheProValG 375
 Db 1288 -----GTTACGTTGCGATC---ATATC 1269
 QY 375 YValTyrTyrCysIleLysGlnIleAsnAspGluArgValPheValValLeuTyrAla 395
 Db 1268 CAGATGGACCTCTGTAGTCCCACTCAATTCGCCGAATCTGTAGTACTTAATTGTACA 1209
 QY 395 eSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuTyrProValVa 415
 Db 1208 TAAGCAGTTGAGCAGGCGCGAGCCCTTCGGCATCTACCTCAATTCACCGCGTCCG 1149
 QY 415 lCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSe 435
 Db 1148 TAAAGTAAATCGCTTTCTTAATGTCTTGGATGCTCTCCCTC----- 1106
 QY 435 rSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSe 455
 Db 1105 -----AGCAATTC 1098
 QY 455 rIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAs 475
 Db 1097 GGACCATTCACAGGAACTGTGTATATCATCCCGCATAGTACCCGATCACACCCCAAGA 1038
 QY 475 pAlaGlnAspThrGlyValSerSerAsnLeuLysSerIleValIleLeuAlaValle 495
 Db 1037 TCACCAAAACGTAGTCCACGTCAGAGATGTCATAATTCGTAGGACTTCTCTCTCGG-TT 979
 QY 495 uMetLeuLeuMetMetPheAlaValHisCysThrTTPValThrSerAsnAlaTyrSerSe 515
 Db 978 GAAGACATTGCCCTGCCAACCGCTATGTGACTATTGTTCCAGTATTATTATCCACT 919
 QY 515 rProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAs 535
 Db 918 AGCGTCGTTCTGTTTGCATTCCTGCTATCT-----GGTATCCGTAATCCACCAAGAC 865
 QY 535 pPheArgGluAlaTyrTyrTyrTyrLeuSer-GlnAsnThrAlaAspAlaArgValMetS 555
 Db 864 ATACCGCGGATCATCGGAGTGTCTCGAAGCCA-----GTAGTAA 820

; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1318

; LENGTH: 560
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-1318

Alignment Scores:
 Pred. No.: 1,21e-101 Length: 560
 Score: 937.00 Matches: 183
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.46% Mismatches: 0
 Query Match: 23.16% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-1318 (1-560)

QY 347 IleAlaSerValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHis 366
 DB 558 ATTGCATCCGTCGGAGCATCAGCCACCACTTGGTTCTTCCTTTGATCTGCAC 499
 QY 367 IleLeuValCysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGlu 386
 DB 498 ATCATGTGTGGCCTTCCAGTGGAGTGTGGTACTGCTCAAGCAGCATCAAGCAGAG 439
 QY 387 ArgValPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 406
 DB 438 CGCGTTTTCGTGTGTGTACGCCATCAGTGGCGTTTACTTCGCTGGTGTGATGGTGGT 379
 QY 407 LeuMetLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeu 426
 DB 378 TTGATGTGTACCCCTCAGCGGTGTGTGTCATGCTGGCCGAGTGGCCTTTTCGGACTG 319
 QY 427 LeuAspValPheLeuGlnLysAspSerSerLysArgMetGlyThrAlaIleSerAlaAla 446
 DB 318 TTGATGTGTTCCTGCAAGAGGATTCTCTAAGCGAATGGGCACAGCCATTAAGCGCAGCC 259
 QY 447 ThrGluValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspIleAlaGly 466
 DB 258 ACCGAAGTGAAGCTGAGGATTCCATTGAGAAGAGACGCTGTACGACAGAGCTGGC 199
 QY 467 LysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeu 486
 DB 198 AAGCTGAAGCATCGTACTAAGCATGTATGCCCGTTCTAATGCTGTGATGATTCGCTCCACTGCACG 139
 QY 487 LysSerIleValIleLeuAlaValLeuMetMetMetPheAlaValHisCysThr 506
 DB 138 AAGAGTATTGTTATTTTGGCCGTTCTAATGCTGTGATGATTCGCTCCACTGCACG 79
 QY 507 TrpValThrSerAsnAlaTyrSerProSerIleValLeuAlaPheHisAsnSerGln 526
 DB 78 TGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTTCCACACAGTCAA 19
 QY 527 AspGlySerArg 530
 DB 18 GATGATCCCGC 7

RESULT 9

US-09-270-767-16600/c
 ; Sequence 16600, Application US/09270767
 ; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16600

; LENGTH: 560
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-16600

Alignment Scores:
 Pred. No.: 1,21e-101 Length: 560
 Score: 937.00 Matches: 183
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.46% Mismatches: 0
 Query Match: 23.16% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-8 (1-774) x US-09-270-767-16600 (1-560)

QY 347 IleAlaSerValSerGluHisGlnProThrThrTrpPheSerPhePheAspLeuHis 366
 DB 558 ATTGCATCCGTCGGAGCATCAGCCACCACTTGGTTCTTCCTTTGATCTGCAC 499
 QY 367 IleLeuValCysAlaPheProValGlyValTyrCysIleLysGlnIleAsnAspGlu 386
 DB 498 ATCATGTGTGGCCTTCCAGTGGAGTGTGGTACTGCTCAAGCAGCATCAAGCAGAG 439
 QY 387 ArgValPheValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArg 406
 DB 438 CGCGTTTTCGTGTGTGTACGCCATCAGTGGCGTTTACTTCGCTGGTGTGATGGTGGT 379
 QY 407 LeuMetLeuThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeu 426
 DB 378 TTGATGTGTACCCCTCAGCGGTGTGTGTCATGCTGGCCGAGTGGCCTTTTCGGACTG 319
 QY 427 LeuAspValPheLeuGlnLysAspSerSerLysArgMetGlyThrAlaIleSerAlaAla 446
 DB 318 TTGATGTGTTCCTGCAAGAGGATTCTCTAAGCGAATGGGCACAGCCATTAAGCGCAGCC 259
 QY 447 ThrGluValAspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspIleAlaGly 466
 DB 258 ACCGAAGTGAAGCTGAGGATTCCATTGAGAAGAGACGCTGTACGACAGAGCTGGC 199
 QY 467 LysLeuLysHisArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeu 486
 DB 198 AAGCTGAAGCATCGTACTAAGCATGTATGCCCGTTCTAATGCTGTGATGATTCGCTCCACTGCACG 139
 QY 487 LysSerIleValIleLeuAlaValLeuMetMetMetPheAlaValHisCysThr 506
 DB 138 AAGAGTATTGTTATTTTGGCCGTTCTAATGCTGTGATGATTCGCTCCACTGCACG 79
 QY 507 TrpValThrSerAsnAlaTyrSerProSerIleValLeuAlaPheHisAsnSerGln 526
 DB 78 TGGGTGACCAAGCAATGCTACTCCAGTCCCTCCATTGCTTGGCTTTCCACACAGTCAA 19
 QY 527 AspGlySerArg 530
 DB 18 GATGATCCCGC 7

RESULT 10

US-09-248-796A-3089
 ; Sequence 3089, Application US/09248796A
 ; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
 ; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 3089
 ; LENGTH: 867

RESULT 12

US-09-270-767-12856
 ; Sequence 12856, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09270767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 12856
 ; LENGTH: 914
 ; TYPE: DNA
 ; ORGANISM: *Drosophila melanogaster*
 US-09-270-767-12856

Alignment Scores:
 Pred. No.: 7,399-79 Length: 914
 Score: 748.50 Matches: 138
 Percent Similarity: 79.59% Conservative: 18
 Best Local Similarity: 70.41% Mismatches: 33
 Query Match: 18.50% Indels: 7
 DB: 4 Gaps: 2

US-10-028-384-8 (1-774) x US-09-270-767-12856 (1-914)

QY 534 AspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAspAlaArgVal 553
 DB 5 GATGACTTCGCGAGCGCTACTGCTGCAGATCAACACTCCGAGAGCGCTCGGATA 64
 QY 554 MetSerTrpTrpAspTyrGlyTrpGlnIleAlaGlyMetAlaAsnThrThrLeuVal 573
 DB 65 ATGTCCTGGTGGGACTACGGCTACCGATACGCGCATGCGCAATCGGACCATATTAGTG 124
 QY 574 AspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSerThr 593
 DB 125 GATAACAATACTTGGACACACACATATATCGCGCTGGCCAGCGATGGCTCTTCG 184
 QY 594 GluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIlePhe 613
 DB 185 GAGGAGAAGCGCTACGAGATAATGAGGGAAGTGGATGGAGTACGATCTCTCGTATTTTC 244
 QY 614 GlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrpMetValArgIle 633
 DB 245 GGAGGCTCAGTCTGCTACTATCGGACGATATCAACAAGTTCCTGTGGATGGTGGCATT 304
 QY 634 Ala-----GluGlyGluHisProLysAspIleLysGluSerAspTyrPheThr 649
 DB 305 GCGCGGACGACGATCGTGTGGCGAC-----ATCCGGAAAGGACTACTATGCG 355
 QY 650 AspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMet 669
 DB 356 GCCAACGGAGTTCGGAGTGGACAGAGGGCTCACCCACTCTCAATTGTTGATG 415
 QY 670 TyrLysLeuSerTyrTrpArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGly 689
 DB 416 TACAAGATGTGCTACTATCGCTTGGGCAAAATGTACACGGAAGGTGGCAAGCCAGGCGC 475
 QY 690 TyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGlu 709
 DB 476 TACGATCGAGTTCGTGGCGCGAGATCGGCAACAGGACTTGAATCGATGTCCTGGAG 535
 QY 710 GluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValLys 725
 DB 536 GAGCGGTACACCGAGAGACTGCTGCTGGTGGCGATCTCAAGGTTAAG 583

RESULT 13

US-09-248-796A-3090
 ; Sequence 3090, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Endege, Wilson O.

APPLICANT: Keith Weinstock et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
 ; FILE REFERENCE: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107195.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 3090
 ; LENGTH: 594
 ; TYPE: DNA
 ; ORGANISM: *Candida albicans*
 US-09-248-796A-3090

Alignment Scores:
 Pred. No.: 3,589-56 Length: 594
 Score: 555.00 Matches: 110
 Percent Similarity: 72.83% Conservative: 24
 Best Local Similarity: 59.78% Mismatches: 38
 Query Match: 13.72% Indels: 12
 DB: 4 Gaps: 5

US-10-028-384-8 (1-774) x US-09-248-796A-3090 (1-594)

QY 11 LysValAlaGlyTyrSer---SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
 DB 52 AAAGTATTGGCGATAGATGTTGAACCTGATTAGAGTTTATTAAAGTTATTATTATATA 111
 QY 30 -----AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIle 45
 DB 112 TCGATAGACAGGTGAGCTATTTCCTTCGTTTATTTCGCGATTGATTCGATTGAAAGTATT 171
 QY 46 IleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGly 65
 DB 172 ATTCATGAATTCGATCCTTGCTTCAATTCGAGCAACCAATATTATTAGTCACTCATTC 231
 QY 66 TrpTyrAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleVal 85
 DB 232 TTTTATGAATTTTGAATTTGATTGATGATAGACTGGTACCCATTGGGAGAGTCACT 291
 QY 86 GlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHis 105
 DB 292 GGTGTACTTTATATCCGGTTTAATGCTGACTTCAGTGCCATT---TGG-----CAT 342
 QY 106 ValLeuAsn-----IleProValHisIleArgAspIleCysValPheLeuAla 121
 DB 343 ATTTTACGTGATTGGTTGCTTACCGCTTGATATTATAAATTGTTGTTTATTAGCA 402
 QY 122 ProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrpSer 141
 DB 403 CCAGTTTTCGGGATTAACCTGCAATTTGACTTATTATTTTGGACTAAAGAAATGAACGAT 462
 QY 142 AlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArg 161
 DB 463 TCTAGTGCACGATTATTGGCAGCTATTATTATGGGATTGCCCGAGGTATATTTCAGA 522
 QY 162 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
 DB 523 TCAGTGGCTGGTTCCTTATGATATAATGAAGCAATTCGCCATTACTTTTAAATGGCAACATTT 582
 QY 182 PheLeuTrpVal 185
 DB 583 TATTCTGGATT 594

RESULT 14

US-09-328-111-332
 ; Sequence 332, Application US/09328111
 ; Patent No. 6262333
 ; GENERAL INFORMATION:
 ; APPLICANT: Endege, Wilson O.

```

/ APPLICANT: Steinmann, Kathleen E.
/ APPLICANT: Aatie, Jon H.
/ APPLICANT: Burgess, Christopher C.
/ APPLICANT: Bushnell, Steven E.
/ APPLICANT: Carroll III, Eddie
/ APPLICANT: Catino, Theodore J.
/ APPLICANT: Dertl, Annan
/ APPLICANT: Ford, Donna M.
/ APPLICANT: Lewis, Marcia E.
/ APPLICANT: Monahan, John E.
/ APPLICANT: Schlegel, Robert
/ TITLE OF INVENTION: NOVEL HUMAN GEN
/ TITLE OF INVENTION: PRODUCTS
/ FILE REFERENCE: CCD-257 (US)
/ CURRENT APPLICATION NUMBER: US/09/09
/ CURRENT FILING DATE: 1999-06-08
/ EARLIER APPLICATION NUMBER: US 60/
/ EARLIER FILING DATE: 1998-06-10
/ NUMBER OF SEQ ID NOS: 850
/ SOFTWARE: fastseq for Windows Ver
/ SEQ ID NO 332
/ LENGTH: 616
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(616)
/ OTHER INFORMATION: n = A,T,C or G
US-09-328-111-332

```

US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

Plant
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 15, 2004, 06:02:06 ; Search time 974.477 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-8
Perfect score: 4046
Sequence: 1 MNRTPKVLNSKVAGYSSLLT.....RRKGYIRNPVVVKGKRLK 774

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10028384/runat_14122004_131523_8140/app_query.fasta_1.3740
-DB=Published Applications NA -QWTF=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEL=0 -COOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10028384@cgn_1_1_2577@runat_14122004_131523_8140
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DBG TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -YgaPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09D_PUBCOMB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq*
- 20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 4046 | 100.0 | 2417 | 15 | US-10-028-384-7 |
| 2 | 3949 | 72.9 | 2484 | 15 | US-10-028-384-1 |
| 3 | 3948 | 72.9 | 2710 | 15 | US-10-028-384-3 |
| 4 | 2158.5 | 54.3 | 3093 | 18 | US-10-417-375-92 |
| 5 | 2158.5 | 54.3 | 3094 | 15 | US-10-028-384-9 |
| 6 | 2158.5 | 54.3 | 3046 | 18 | US-10-417-375-95 |
| 7 | 2158.5 | 54.0 | 5404 | 18 | US-10-417-375-99 |
| 8 | 2158.5 | 54.0 | 5827 | 18 | US-10-417-375-97 |
| 9 | 2177.5 | 53.8 | 2472 | 15 | US-10-171-581-112 |
| 10 | 2177.5 | 53.8 | 2472 | 15 | US-10-028-384-11 |
| 11 | 2177.5 | 53.8 | 2472 | 15 | US-10-172-118-742 |
| 12 | 2177.5 | 53.8 | 2472 | 16 | US-10-342-887-742 |
| 13 | 2176 | 53.7 | 2256 | 15 | US-10-032-585-6323 |
| 14 | 2172.5 | 53.7 | 2232 | 15 | US-10-128-714-7139 |
| 15 | 2113.5 | 52.2 | 2466 | 16 | US-10-320-797-2305 |
| 16 | 2068 | 51.1 | 2681 | 16 | US-10-425-114-14408 |
| 17 | 2060.5 | 50.9 | 2779 | 17 | US-10-437-963-99904 |
| 18 | 2059.5 | 50.9 | 2839 | 18 | US-10-425-115-150745 |
| 19 | 2036.5 | 50.3 | 2157 | 9 | US-09-801-368-387 |
| 20 | 2036.5 | 50.3 | 2157 | 18 | US-10-793-639-318 |
| 21 | 2027.5 | 50.1 | 2733 | 15 | US-10-028-384-5 |
| 22 | 2014 | 49.8 | 2603 | 15 | US-10-128-714-6139 |
| 23 | 2014 | 49.8 | 3969 | 15 | US-10-128-714-139 |
| 24 | 2014 | 49.8 | 4603 | 15 | US-10-128-714-5139 |
| 25 | 1888.5 | 46.7 | 1848 | 15 | US-10-128-714-2139 |
| 26 | 1861.5 | 46.0 | 3141 | 18 | US-10-425-115-130787 |
| 27 | 1853 | 45.8 | 3197 | 16 | US-10-424-599-111541 |
| 28 | 1851.5 | 45.8 | 2882 | 16 | US-10-320-797-1305 |
| 29 | 1851.5 | 45.9 | 4738 | 15 | US-10-320-797-305 |
| 30 | 1840.5 | 45.5 | 1969 | 15 | US-10-128-714-1139 |
| 31 | 1608.5 | 39.8 | 2244 | 17 | US-10-437-963-48342 |
| 32 | 1512 | 37.4 | 1543 | 10 | US-09-974-879-133 |
| 33 | 1512 | 37.4 | 1543 | 10 | US-09-305-736-133 |
| 34 | 1512 | 37.4 | 1543 | 10 | US-09-818-683-133 |
| 35 | 1512 | 37.4 | 1543 | 11 | US-09-818-683-133 |
| 36 | 1512 | 37.4 | 1543 | 16 | US-10-621-401-133 |
| 37 | 1509.5 | 37.3 | 1209 | 15 | US-10-106-698-330 |
| 38 | 1509.5 | 37.3 | 1209 | 16 | US-10-264-237-412 |
| 39 | 1282 | 31.7 | 1828 | 10 | US-09-945-527-62 |
| 40 | 1231 | 30.4 | 1728 | 16 | US-10-424-599-122476 |
| 41 | 1147.5 | 28.4 | 2660 | 16 | US-10-264-049-630 |
| 42 | 1006.5 | 24.9 | 1114 | 16 | US-10-296-115-629 |
| 43 | 973 | 24.0 | 1094 | 17 | US-10-437-963-99902 |
| 44 | 750.5 | 18.5 | 887 | 16 | US-10-424-599-77697 |
| 45 | 727 | 18.0 | 500 | 9 | US-09-998-598-1643 |

ALIGNMENTS

RESULT 1
US-10-028-384-7
; Sequence 7, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1.
; SEQ ID NO 7
; LENGTH: 2417
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF132552
; DATABASE ENTRY DATE: 1999-04-27

RELEVANT RESIDUES: (1)...(2417)
US-10-028-384-7

Alignment Scores:

Pred. No.: 0 Length: 2417
Score: 4046.00 Matches: 774
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-8 (1-774) x US-10-028-384-7 (1-2417)

QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuLeuThr 20
DB 71 ATGAATCGGACGCGGAGATGCTGAACACAGAGGTGGCTACAGAGCTTAATCACC 130
QY 21 PheAlaIleLeuLeuLeuAlaTTPLeuAlaGlyPheSerSerArgLeuPheAlaValle 40
DB 131 TTGGCCATCTCGTAAATCGCTGGCTGGCGGGAATTTCTCTCGCTCTTGGCGTATC 190
QY 41 ArgPheGluSerIleIleHisGluPheAspProTTPPheAsnTyrArgAlaThrAlaTyr 60
DB 191 CGTTTCGAGTCGATTATCCATGAGTTTGATCCGCTGTTCAACTACCGGGCCACCGCTAC 250
QY 61 MetValGlnAsnGlyTTPyrAsnPheLeuAsnTTPPheAspGluArgAlaTTPyrPro 80
DB 251 ATGGTCGAGATGGTTGTACAACTTCCTCACTGGTTCGACGAGCGCATGTATCG 310
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIle 100
DB 311 CTCGGCAGGATTTGGCGGCTACCGTCTATCCGCGCTGATGATTACGTCGGCGGAATC 370
QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
DB 371 CATTTGGCTGTGCACGCTACTCAACATACCGGTTCCATATTCGTGACATCTGGTCTCTG 430
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuLeuThrLysGluLeuTrp 140
DB 431 GCGCGGATCTTCACTGGCTGACCTCCATCTCCACCTACCTGCTGCTGACCAAGAGCTGG 490
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
DB 491 TCCGCGGCGCGGCTCTTCGCGCGCAGCTTCATCGCATCGTGGCTGGCTATCATAGT 550
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
DB 551 AGSTCGGTGGCTGGATCGTACGATACAGAGGCGATTTGCCATATTCGCGCTCGAGTTCA 610
QY 181 TyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSerAlaAlaAlaAla 200
DB 611 TACTTCTGTTGGGTGGCTCAGTGAAGACTGGATCGGTGTTCTGGTGGCGCGAGCGCT 670
QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220
DB 671 TTGTCTTACTTCTACATGATGTCGCGCTGGGTGGCTACGTTGCTATCATCAACTGATA 730
QY 221 ProLeuHisValPheValLeuLeuIleMetGlyA-gTyrSerProArgLeuLeuThrSer 240
DB 731 CCCCTGCAGCTCTTCGTACTGCTCATTAATGGCAGGTACTCGCGCGTCTCTGCACCGC 790
QY 241 TyrSerThrPheTyrIleLeuGlyLeuPheSerMetGlnIleProPheValGlyPhe 260
DB 791 TACAGACCTTCTACATCTGGGACTGCTGTTCTCCATGCGATCCCTCTGGGGATTC 850
QY 261 GlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAla 280
DB 851 CAACCGATACGACACGAGTGAACACATGCTGGCTGGGAGTGTGTGCTCTCTTATGGCC 910
QY 281 ValAlaThrLeuArgHisIleGlnSerValLeuSerArgAsnGluPheArgLysLeuPhe 300
DB 911 GTGGCCACCTTTCGCCATTTCGAGTCCGCTGTCTCGCGAAGAGTTCGGGAAGCTGTTC 970

QY 301 IleValGlyGlyLeuLeuValGlyValGlyValPheValAlaValValValLeuThrMet 320
DB 971 ATCGTTCGCGCGGATTTGCTGGTGGCGTGGCGTCTTTTGGCCGCTCGTGGTCTACCATG 1030
QY 321 LeuGlyValValAlaProTTPSerGlyArgPheTyrSerLeuTrpAspThrGlyTyrAla 340
DB 1031 CTGGCGGTGTGGCGCGGTGGAGTGGAGCTTCTACTCGCTGGGATACTGGCTACGCC 1090
QY 341 LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer 360
DB 1091 AAGATCCACATTCCTCATTCATTCGTCGAGCATCCAGCCACCATTTGGTTCCTCG 1150
QY 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTrpTyrCysIle 380
DB 1151 TTCTTCTTTGATCTGCACATCTCTGGTGGCGCTTCCAGTGGGAGTGGTACTGCATC 1210
QY 381 LysGlnIleAsnAspGluArgValPheValValLeuTyrAlaIleSerAlaValTyrPhe 400
DB 1211 AAGCAGATCAACGACGAGCGGCTTTCTGGTGTCTGTACGCCATCAGTGGGTACTTC 1270
QY 401 AlaGlyValMetValArgLeuMetLeuThrIleuThrProValValCysMetLeuAlaGly 420
DB 1271 GCTGTGTGATGGTGGCGCTTTGATGTTGACCTCACGCGGTGGTGTGCATGTGGCCGA 1330
QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
DB 1331 GTGGCTTTTCGGGACTGTTGGATGTCTCTGCAAGAGGATTCGTCTAAGCGAATGGGC 1390
QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGlnLysIleThr 460
DB 1391 ACAGCATTAAGCGACGCGCACCGAAGTGGATGAAGCTGAGGATTCATTCAGAGAGAGAG 1450
QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisIleArgThrLysHisAspAlaGlnGlnAspThr 480
DB 1451 CTGTACGACAAAGCTGGCNAAGCTGAAGCATCGTACTAAGCATGATCCCCAGCAGGATACT 1510
QY 481 GlyValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMetMet 500
DB 1511 GCGGTGACCTCCAACTGGAAGAGATTTGTTATTTTGGCGTCTTAATGCTGTGTGATG 1570
QY 501 PheAlaValHisCysThrTTPValThrSerAsnAlaTyrSerSerProSerIleValLeu 520
DB 1571 TTGCTGTCCACTGCTCACGTGGTGACCAAGTCCCTACTCCAGTCCCTCCATTTGCTTG 1630
QY 521 AlaPheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyr 540
DB 1631 GCTTTCCAAACAGTCAAGATGGATCCGCAACATTTTAGACGATTCAGAGAGGCTTAC 1690
QY 541 TyrTrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTrpAspTyrGly 560
DB 1691 TACTGGCTTTCCGACAGACACTGCGCATGATGCTCGGTTATGCTTTGGTGGGATACGGA 1750
QY 561 TyrGlnIleAlaGlyMetAlaAsnArgThrLeuValAspAsnAsnThrTrpAsnAsn 580
DB 1751 TACCAGATAGCGGAATGGCAACAGAACACGCTAGTGGATTAATAATACGTGGAAACAAT 1810
QY 581 SerHisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIle 600
DB 1811 AGTCACATAGCGCTGGTTGGCAAGCAATGCTTCAACGAGGAGAAAGTCTTACGAAAT 1870
QY 601 MetThrSerLeuAspValLeuValLeuValIlePheGlyValIleGlyTyrSer 620
DB 1871 ATGACATCTCTGACGTGGACTACGTTTGGTGTATCTTTGGCGGTGTGATCGGCTATTCT 1930
QY 621 GlyAspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLys 640
DB 1931 GCGCATGATATCAACAAGTTCCTGTGGATGGTCCGAATTCCTCAACGAGGAGAGCATCCCAAG 1990
QY 641 AspIleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGly 660
DB 1991 GACATTAGGAAGACGATTAATTTACCGACCGCGGTGAATTCAGCGTAGATCCGAGAGT 2050
QY 661 AlaProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeu 680

```

Db 2051 GCTCCGGCCCTGCTCAACTGCTTATGTACAAATTAAGCTPACTACAGATTCGGGGAATTG 2110
Qy LysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsn 700
Db 2111 AAGTTGGACTACAGAGTCCATCTGGATATGATCGCACACGTAAACCGCTCATGGGAAT 2170
Qy LysAspPheAspLeuThrTyrLeuGluAlaValTyrThrGluHisTrpLeuValArg 720
Db 2171 AAGGACTTCGATCTGACCTACCTGGAGAGGCTTCACACAGAACACTGGCTTGTTCGC 2230
Qy IleTyrArgValLysLysProHisGluPheAsnArgProSerLeuLysThrLysGluArg 740
Db 2231 ATCTATAGGGTGAAGAAGCCGATGAGTCAATAGACCATCACTGAAGACCAAGGAGAGA 2290
Qy ThrIleProProAlaAsnPheIleSerArgLysAsnSerLysArgLysGlyTyrIle 760
Db 2291 ACGATTCTCCAGCAAACTTCATTTCGAGAAAGAACTCTAAGCGTCGCAAGGCTACATA 2350
Qy ArgAsnArgProValValValLysGlyLysArgThrLeuLys 774
Db 2351 CGAAACCGACCGGTGTTGTTAAGCGAAACGAACCTTGAAA 2392

```

RESULT 2

```

US-10-028-384-1
; Sequence 1, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 2481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2481)
; OTHER INFORMATION:
US-10-028-384-1

```

```

Alignment Scores:
Pred. No.: 1,74e-293 Length: 2481
Score: 2949.00 Matches: 556
Percent Similarity: 83.88% Conservative: 89
Best Local Similarity: 72.30% Mismatches: 112
Query Match: 72.89% Indels: 12
DB: 15 Gaps: 7

```

US-10-028-384-8 (1-774) x US-10-028-384-1 (1-2481)

```

Qy 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeu 29
Db 187 TCGAGCGCGGTGGGTGGAGTGGCTTCTCTCTCCATCCATCTCTCTGGCGCTGGCTT 246
Qy 30 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
Db 247 GCGGCTTCAGCTCGCGCTCTTCGCGCTCATCGCTTCGAAAGCAATCATCCAGAGTTC 306
Qy 50 AspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
Db 307 GACCGGTGTTTAACTATAGATCAACATCATCTGTCATCTCATCGGTCTATGAATTT 366
Qy 70 LeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArgIleValGlyGlyThrVal 89
Db 367 TTTAATGCTTGTGTAAGAGCGAGTGTATCCACTAGAGAGATAGTAGTGTACTGTT 426

```

```

Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle 109
Db 427 TACCCAGGGTGTATGATAAACCCTGGCTTATTCATTTGGATTTAAATACATTAACATA 486
Qy 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
Db 487 ACTGTTCCATATAGAGACGTATGTGTCTTCCTTGCACCACTTTTAGCGGCTTACACT 546
Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaIle 149
Db 547 ATATCTACTTTCTGCTTACAGAGAACTTGGAAACCAAGAGGAGGAGACTTTTAGCTGCT 606
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
Db 607 TGTTTATTCCTATTGACAGGCTACATATCTCGTCACTAGCTGATCTCTTGTATAT 666
Qy 170 GluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLys 189
Db 667 GAAGGCACTGCTATTTTGCATTCAGTTTACATATTTATTCACATCATGATTTGTGTGTT 726
Qy 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 727 ACTGGGTCAGTTTGGACAATGTCTGCTCTATCTCTATATATGCTCTCTGCT 786
Qy 210 TrpGlyGlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIle 229
Db 787 TGGGGTGGTTATGATTTATTCATCAATATTTATTCACATCATGATTTGTGTGTTTACTG 846
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
Db 847 ATGCAGAGATACAGCAAAAGAGCTCACATAGCATATAGCACATTTCTACATTTGGGTTA 906
Qy 250 LeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 907 ATATATCAATGCAGATACCTTTTGGGATTTCCAGCCAAATCAGAACAGTGAACACATG 966
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 967 GCAGCTGCAGGTGCTTTGCAATGTCTGCAAGCTTATGCTTTCTGCAGTATCTGAGAGAC 1026
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuValGlyVal 309
Db 1027 CGATTAAACAAACAGAGATTCCAGACCTTTTCTTTTGGGTGTATCATCTGCTGAGTGC 1086
Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGly 329
Db 1087 GCTGTGTCCTTAGTGTCACTATTGACTTATACAGGTATCATTCACCATGGAGTGGC 1146
Qy 330 ArgPheTyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSer 349
Db 1147 AGGTTTATTCATTTGGGATACCTGGGTATGCAAAATATACATTCATTCATTCATCA 1206
Qy 350 ValSerGluHisGlnProThrThrTyrPheSerPhePheAspLeuHisIleLeuVal 369
Db 1207 GTGCTGAGCATCACTACGACTTGGGTGCTCTTCTTCTTGTGATCTACATATCTGTA 1266
Qy 370 CysAlaPheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1267 TGTACCTTCCAGCAGCGCTTGGTTCCTGCATCAAAATATCAACGATGAAAGATATT 1326
Qy 390 ValValLeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeu 409
Db 1327 GTTGTCTATATGCATCATGCTGCTCTACTTTGCTGGAGTGTGTTGGTGGCTGATGTG 1386
Qy 410 ThrLeuThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuAspVal 429
Db 1387 ACTTTGACTCCAGTCGTGTATGCTGTGCTCAATTTGCTTTTCAAAATTTGTTTGGACAC 1446
Qy 430 PheLeuGlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1447 TATTTGGGATGAC---ATGAAAGGGAATAATCCACCTGTGGAGACGACGATGATGAG 1503
Qy 450 AspGluAlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeu--- 468

```

```
Db 1504 GAT-----GACAAAAGAACCAAGGAATTTGTATGATAGGCGAGGTAAAGTGAGG 1554
Qy 469 LyHisArgThrIysHisAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSer 488
Db 1555 AAACATGCAACTGAACAGGAAAAAATGAAGAG--GGATTAGGCGCCCTAATATATAAAGC 1611
Qy 489 IleValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTrpVal 508
Db 1612 ATTGTCAACATGTTGATGCTGATGCTATTGATGATGTTTGTGTCCACTGTACCTGGGTC 1671
Qy 509 ThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGly 528
Db 1672 ACAAGCAATGCCCTACTCTAGTCAAGTGTAGTCCCTGGCCTCATACAAT---CATGATGGC 1728
Qy 529 SerArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThra 548
Db 1729 ACCAGGAATATCTTAGATGATTTAGAGAGCTTACTTTGGCTAAGCCAAAATACAGAT 1788
Qy 549 AspAspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyWeAlaAsn 568
Db 1789 GAACATGCGACGAGTAATGCTTGGTGGGATTATGGCTATCATGATAGCTGGAAATGCTAAT 1848
Qy 569 ArgThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLys 588
Db 1849 AGAATCTACGTTGGTGATTAATACACCTGGATATACACCCATAGCATGCTGGTGGAAA 1908
Qy 589 AlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyr 608
Db 1909 GCTATGCTCTTAATGAACACAGCAGCTATAAAATCATGAGGACTCTAGATGTAGATTAT 1968
Qy 609 ValLeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeu 628
Db 1969 GTTTTGGTATTTTGGAGGGTTATTTGGCTATTCTGTGATGATATCAAAATTTCTC 2028
Qy 629 TrpMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
Db 2029 TGGATGTTAGGATAGCTGAAGAGAACTATCCCAAGACATTCGGGAAAGTGACTATTAT 2088
Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyValProAlaLeuAsnCysLeu 668
Db 2089 ACCCCACAGGAGAAATTCGTTAGACAAAGCAGATCCCTACTTTCTGAAATGGCTT 2148
Qy 669 MetTyrIysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2149 ATGTATAAATGTCACTACTACAGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCA 2208
Qy 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
Db 2209 GATTTTGACCGACACGTAATCTGTGATTTGGAATAGGACATTAATTTCAACATTTG 2268
Qy 709 GluGluAlaTyrThrGluHisTrpLeuValArgIleTyrArgValIysPheHis 728
Db 2269 GAAGAAAGCCTTTACATCAGAACACTGCTTGTAGGATATATAAGTAAAGACACCTGAT 2328
Qy 729 GluPheAsnArgProSerLeuLysThrLysGluArgThr-----IleProProAla 745
Db 2329 -----ACAGGAGAGACATTAGATCAAAACCTCGAGTCACCAACATTTCCCAAAACAG 2382
Qy 746 AsnPheIleSerArgIysAsnSerIysArgArgLysGlyTyrIleArgAsnArgProVal 765
Db 2383 AAGTATTGTCAAAGAAAGACTACCAAAAGAAAGCGTGGCTACATTAATAAATAAGCTGGTT 2442
Qy 766 ValValLysGlyLysArgThrLeuLys 774
Db 2443 TTTAAGAAAGCGCAAGAAATATCTAAG 2469
```

RESULT 3

US-10-028-384-3

; Sequence 3, Application US/10028384

; Publication No. US20030148285A1

; GENERAL INFORMATION:

; APPLICANT: COMPATIGENE INC.

```
APPLICANT: PERREAULT, Claude
APPLICANT: MCBIDE, Kevin
TITLE OF INVENTION: Mammalian STMP, gene sequence and uses thereof in cancer therapy
FILE REFERENCE: 5600-74
CURRENT APPLICATION NUMBER: US/10/028,384
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn, version 3.1
SEQ ID NO 3
LENGTH: 2710
TYPE: DNA
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: AK018758
DATABASE ENTRY DATE: 2001-07-05
RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3
```

Alignment Scores:

```
Pred. No.: 2,53e-293 Length: 2710
Score: 2948.00 Matches: 551
Percent Similarity: 83.14% Conservative: 90
Best Local Similarity: 71.47% Mismatches: 114
Query Match: 72.86% Indels: 16
DB: 15 Gaps: 5
```

US-10-028-384-8 (1-774) x US-10-028-384-3 (1-2710)

```
Qy 10 SerLysValAlaGlyTyrSerSerLeuIleThrPheAlaIleLeuLeuLeuLeuLeu 29
Db 254 TGCACGCGCGCGGGTGGCGAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 313
Qy 30 AlaGlyPheSerSerArgLeuPheAlaValIleArgPheGluSerIleIleHisGluPhe 49
Db 314 GCCGCTTCACTCGCGCTCTTCCGCTCATCCGCTCGAGAGCATCATCCACAGATTC 373
Qy 50 AspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPhe 69
Db 374 GACCCGCTGTTTAACTATAGATCAACACATCATCTCTGTCATCTCATGATTTCTATGATTT 433
Qy 70 LeuAsnTrpPheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGlyThrVal 89
Db 434 CTAAATTTGTTTGTGAAGAGCATGTATCCACTGGGAGAGATAGTGGGTGCACCGTT 493
Qy 90 TyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIle 109
Db 494 TACCCAGCGTTTGTATGAACAGCTGGCGCTTATTCTTGGATTTTAAATACATTGAACATA 553
Qy 110 ProValHisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSer 129
Db 554 ACAGTTTCAATAGAGATGTGTGTATCTCTTGCCACCACTTTTAGCGGCTTATATACC 613
Qy 130 IleSerThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAla 149
Db 614 ATATCTAGCTTCTGCTAACTAGAGAACTGTGGAACCAAGCAGCAGCTTCTAGCTGCT 673
Qy 150 SerPheIleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 169
Db 674 TGTTCATTTGTATCTACAGGATATATCTCGGTACGTGGCGGATCTTTGATAT 733
Qy 170 GluGlyIleAlaIlePheAlaLeuGluPheThrTyrPheLeuTrpValArgSerValLys 189
Db 734 GAAGGCAATGGCATTTTGGCGCTTCACTTCACTTACTTATGGGTAAAGCTGTGAAG 793
Qy 190 ThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAla 209
Db 794 ACCGGCTCTGTCTCGACATATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 853
Qy 210 TrpGlyGlyTyrValPheIleAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 229
Db 854 TGGGAGGATATGTGTCTCATCATCAACCTCATCCCTCTCCATGTGTGTGTGTGTGTGTG 913
Qy 230 MetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeu 249
```

Db 914 ATGCAGAGTACAGACAGAGAGTCTACATAGCATATAGCACTTTGACATTTGGGGTTTA 973
Qy 250 LeupheSerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMet 269
Db 974 ATATTATCCATGACATACCTTTTGTGGATTTTCAGCCATCAGAACAGCGAGCACATG 1033
Qy 270 AlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSer 289
Db 1034 GCAGCTGCAGGTGCTTTGCGCTCTCAAGCTTACGCTTTTGTGAGTATCTGAGAGAC 1093
Qy 290 ValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyVal 309
Db 1094 CGGTGACAAACAGGAGCTCCAGACCCTTTCTTTTGGGTGCTCTCACTAGCTGCAGGC 1153
Qy 310 GlyValPheValAlaValValLeuThrMetLeuGlyValValAlaProTTPSerGly 329
Db 1154 GCTGTGTTCTCTAGTGTCTCTCTGACATACACAGTTATATGACCATGAGGTGGC 1213
Qy 330 ArgPheTyrSerLeuTTPAspThrGlyTyrAlaLysIleHisIleProIleAlaSer 349
Db 1214 AGGTTTATTCATATGGATCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1273
Qy 350 ValSerGluHisGlnProThrTTPPheSerPhePhePhePhePhePhePhePhePhe 369
Db 1274 GTGCTGAACATCAGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1333
Qy 370 CysAlaPheProValGlyValTTPCysIleLysGlnIleAsnAspGluArgValPhe 389
Db 1334 TGTACCTTCCAGCAGGCTATGCTTCTGATCAAAATATCAACGATGAAGATATTT 1393
Qy 390 ValValLeuTTPAlaIleSerAlaValTTPPheAlaGlyValMetValLeuMetLeu 409
Db 1394 GTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1453
Qy 410 ThrLeuThrProValValCysMetLeuAlaValAlaPheSerGlyLeuLeuAspVal 429
Db 1454 ACTCTGACCCGGTCTGCTGATGCTGCTGCGCATGCTTCTTCCATGTTTGTGACAC 1513
Qy 430 PheLeuGlnIleAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluVal 449
Db 1514 TATTTGGGGATGATGATGAAGGAA-----AACCCACCTGTG 1552
Qy 450 AspGluAlaGluAspSerIleGluLysLys-----ThrLeuTTPAspLysAlaGly 466
Db 1553 GAGGACAGCAGTGTAGGATGACAAAGAAACCCAGGAACTTGTATGACAAAGCAGGT 1612
Qy 467 LysLeuLysHisArgThrLysHisAspAlaGlnAspThrGlyValSerSerAsnLeu 486
Db 1613 AAAGTGAGAGCATGTGACAGACAAAGAACCTGAGAGGGCTTGGGCCCAACATC 1672
Qy 487 LysSerIleValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThr 506
Db 1673 AAAAGCATTTGACCATGCTGATGCTCATGCTCTGATGATGATGCTGCTGCTGCTGCTG 1732
Qy 507 TTPValThrSerAsnAlaTTPSerProSerIleValLeuAlaPheHisAsnSerGln 526
Db 1733 TGGGTCAACAGCAGCCTATCTCAGTCCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
Qy 527 AspGlySerArgAsnIleLeuAspAspPheArgGluAlaTTPTTPLeuSerGlnAsn 546
Db 1790 GATGTTACAGGATATATATAGATGATTTTAGAGAGCGTACTTTTGGCTGAGACAAAC 1849
Qy 547 ThrAlaAspAlaArgValMetSerTTPAspTTPArgTTPGlnIleAlaGlyMet 566
Db 1850 ACGGATGAACACCCCGGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1909
Qy 567 AlaAsnArgThrThrLeuValAspAsnAsnThrTTPAsnAsnSerHisIleAlaVal 586
Db 1910 GCCAACAGGACCACTCTGTTGGTAAACAAACACCTGGAACAAACACCAACCAACCAACCA 1969
Qy 587 GlyLysAlaMetSerSerThrGluGluLysSerTTPClnIleMetThrSerLeuAspVal 606

Db 1970 GGAAACGCTATGCTTCCCAATGAACCGCGGCTATATAAATCATGAGGTCCCTTGTATGTC 2029
Qy 607 AspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLys 626
Db 2030 GATTTATGTTGTTATTTTCGGAGGAGTGAITGGTATTCGGGGAGCATATCAACAAG 2089
Qy 627 PheLeuTTPMetValArgIleAlaGluGlyGluHisProLysAspIleLysSerAsp 646
Db 2090 TTTCTCTGGATGTGAGTAGCTGAAGGGAGCATCCCAAGACATCCGGAGAGGTGAC 2149
Qy 647 TYPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsn 666
Db 2150 TATTTCCACCCAGCAGGAGAGTCCGAGTAGACAAAGCTGGGTCTCTACTCTGTATAAC 2209
Qy 667 CysLeuMetTYPLeuLysLeuSerTYPArgPheGlyGluLeuLysLeuAspTYPArgGly 686
Db 2210 TGCCTTAATGATATAAATGTCATCTACAGATTGGAGAAATCAGTAGATTTTCGACT 2269
Qy 687 ProSerGlyTYPAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThr 706
Db 2270 CCCCCAGGCTTTGACCAACACGTAATGCTGAGATTGGAATAAAGACATTAAATTCAG 2329
Qy 707 TyrLeuGluGluAlaTYPThrThrGluHisTTPLeuValArgIleTYPArgValLysLys 726
Db 2330 CATTTGGAGGAGCTTTTACATCGAGCACTGGCTTGTGAGATATATAAATGAAGCA 2389
Qy 727 ProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----IlePro 743
Db 2390 CCTGAC-----AACAGGGAGACATAGGTTCACAAACCTCGAGTCACCAACATCGTCCC 2443
Qy 744 ProAlaAsnPheIleSerArgLysAsnSerLysArgGlyTYPLeuArgAsnArg 763
Db 2444 AACACAGAGTATTTGTCAAAGAACACTACTATAAAGAGCGGTGCTTACGTTAAAAATAAG 2503
Qy 764 ProValValLysGlyLysArgThrLeuLys 774
Db 2504 CTAGTGTTTAAGAAAGCAAGAGACCTCTAAG 2536

RESULT 4
US-10-417-375-92
; Sequence 92, Application US/104:7375
; Publication NO. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 3093
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-417-375-92

Alignment Scores:
Pred. NO.: 6,92e-216 Length: 3093
Score: 2198.50 Matches: 416
Percent Similarity: 73.78% Conservative: 113
Best Local Similarity: 58.02% Mismatches: 143
Query Match: 54.34% Indels: 45
DB: 18 Gaps: 8

US-10-028-384-8 (1-774) x US-10-417-375-92 (1-3093)

Qy 17 SerLeuLeuThrPheAlaIleLeuLeuIleAlaTTPLeuAlaGlyPheSerArgLeu 36
Db 157 ACATCTCAAGCTTCTCATCTGTCGATGGCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCT 216
Qy 37 PheAlaValIleArgPheGluSerIleHisGluPheAspProTTPPheAsnTYPArg 56

Db 217 TTCTGCTGCTGAGATTGAAAGTGCATCCATGAGTTGATCCGCTACTTAAATATATCGG 276
 Qy 57 AlaThrAlaTyrMetValGlnHsnGlyTyrPheAsnPheLeuAsnTyrPheAspGluArg 76
 Db 277 ACTACCCGGTTTCTGGCTGAGGAGGGTTTATAAATTCCTAATCCAACTGCTTTGATGACCGG 336
 Qy 77 AlaTyrPheProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
 Db 337 GCTTGTACCTTTGGCCGCAATCATGGAGGACAAATTACCCAGGTTTAAATGATCACT 396
 Qy 97 SerGlyGlyIleHisTyrPheLeuHisValLeuAsnIleProValHisIleArgAspIle 116
 Db 397 TCTGCTCAATCTACCATCTACTCCATTTCTTCCATCACTCACTATTGACATTCGAATGTC 456
 Qy 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
 Db 457 TGTGTTTCTGCTGCCCACTTTCTCTCTTCCACCATCGTTACGTACCACTTACC 516
 Qy 137 LysGlnLeuTyrPheSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
 Db 517 AAGAGCTCAAGGATGAGGAGTGGGCTTCTTGTCTGCTGCCATGATTCGTGTAGTTCCT 576
 Qy 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlnGlyIleAlaIlePheAla 176
 Db 577 GGTATATTTCTGATCTGATGCTGCTCTATGATGATGATGATGATGATGATGATGATG 636
 Qy 177 LeuGlnPheThrTyrPheLeuTyrPheValArgSerValLysThrGlySerValPheTyrSer 196
 Db 637 ATGCTGCTTACTTACTACATGATGATGATGATGATGATGATGATGATGATGATGATG 696
 Qy 197 AlaAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrGlyGlyTyrValPheIle 216
 Db 697 GCCAAGTGTGCCCTGCTTACTTACTATGATGATGATGATGATGATGATGATGATGATG 756
 Qy 217 IleAsnLeuIleProIleHisValPheValLeuLeuLeuMetGlyArgTyrSerProArg 236
 Db 757 ATCAACTGATCTCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
 Qy 237 LeuLeuThrTyrSerTyrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
 Db 817 ATCTACGTAGCTACTGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
 Qy 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeuGlyValPheVal 276
 Db 877 TTGTTGGTTCAGCCCGCTTCTCATCAGAACATGCGCAGCTTTGGAGTGTGTTGGT 936
 Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 Db 937 CTCTGTCAGATCCATGCTTTCGATGATGATGATGATGATGATGATGATGATGATGATG 996
 Qy 297 ArgLysLeuPhe-----IleValGlyLeuLeuValGlyValGlyVal 311
 Db 997 GAAGTCTTTCGCGAGTGTATCTCCCTGGTGTGCTTGTCTCTCTCTCTCTCTCTCTCTCT 1053
 Qy 312 PheValAlaValValLeuLeuMetLeuGlyValValAlaProTyrPheSerGlyArgPhe 331
 Db 1054 -----GCTCTCTCTCATGCTAAC-----GGAAAAATTTCTCCCTGAGCGGGTTC 1101
 Qy 332 TyrSerLeuTyrPheThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
 Db 1102 TACTCTCTGCTGATCCCTCTTATGCTAAGATAACATTCCTCATTTATGCTATCTGTTCT 1161
 Qy 352 GluHisGlnProThrTyrPheSerPhePheAspLeuHisIleLeuValCysAla 371
 Db 1162 GAGCACCGCCCAACCTGCT 1221
 Qy 372 PheProValGlyValTyrPheTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
 Db 1222 TTTCAGTGTGCTCTTACTCTCTTACCAACCTGTCTGATGCTGCGGATTTTATCATC 1281
 Qy 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuLeuThrLeu 411
 Db 1282 ATGTATGCTGCTGACGATGATCTTTTTCAGCTGTAATGTCGCTGTAATGTCGCTGATG 1341

Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
 Db 1342 GCACCTGTATGTGTCATCTTTTCCGATCTGTTTCCAGGTGCTGCCCATATATG 1401
 Qy 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
 Db 1402 AAAAATCTGGACATAAAGTCGCCA----- 1425
 Qy 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
 Db 1426 -----GACAAAGAGCAAG----- 1440
 Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
 Db 1441 -----AAGCAACAGGATCTACTTACCCTATTAAAGATGAGTGGCGAGTGGG 1488
 Qy 490 ValIleLeuAlaValLeuMetLeuLeuMetPheAlaValHisCysThrTyrValThr 509
 Db 1489 ATGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1548
 Qy 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
 Db 1549 ACTGAAGCTTATCTTCTCCCTCCATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1605
 Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTyrLeuSerGlnAsnThrAlaAsp 549
 Db 1606 AGGATCATTTTGTGATGCTTCCGAGAGCGTATTATTGCTGCTGCTGCTGCTGCTGCTGCTG 1665
 Qy 550 AspAlaArgValMetSerTyrTyrPheTyrPheTyrGlnIleAlaGlyMetAlaAsnArg 569
 Db 1666 GATGCAAAAGTCATGTCATGCTGGGATTAAGCTTACCATAATCTCAATGGCAATCGG 1725
 Qy 570 ThrThrLeuValAspAsnAsnThrTyrPheAsnSerHisIleAlaLeuValGlyLysAla 589
 Db 1726 ACAATTTTGTGAGCAATAACACATGAATAAATCCCATATTTCTCGAGTAGGCGAGCA 1785
 Qy 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
 Db 1786 ATGGCATCCACAGAAAGAAAGCCATGAAATCATGAGGAGCTTGTATGCTAGCTATGTG 1845
 Qy 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspPheLeuLysPheLeuTyr 629
 Db 1846 CTGTGCTTTTGGAGCGCTTACTGCGGTATCTTCTGATGATATCAACAGTTCTTCTGG 1905
 Qy 630 MetValArgIle-----AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
 Db 1906 ATGGTCCGATTTGGAGGAGGACACAGACAGGAGGACACATTAAGGAGAAATGACTACTAT 1965
 Qy 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
 Db 1966 ACTCTACTGCGGAATCCGGTTGTATGCTGAGGAGTTCTCCGCTGCTGCTCAACTGCCCT 2025
 Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
 Db 2026 ATGTACAAATGTGTACTACCGCTTTGGCGAGCTCTACACAGAACCAAGCGTCCACCA 2085
 Qy 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysPheAspLeuThrTyrLeu 708
 Db 2086 GCTTTGTACCGTGTTCGAATGCTGAGATTGGTAATAAGACTTTGAGCTTGATGCTGCTG 2145
 Qy 709 GluGluAlaTyrThrThrGluHisTyrPheValArgIleTyrArgValLys 725
 Db 2146 GAGGAAGCGTATACACAGAACACTGCTAGTCAGGATATACAAAGTAAAG 2196

RESULT 5

US-10-028-384-8
 ; Sequence 9, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCBRIDE, Kevin

/ FILE REFERENCE: 5600-74
 / CURRENT APPLICATION NUMBER: US/10/028.384
 / CURRENT FILING DATE: 2001-12-20
 / NUMBER OF SEQ ID NOS: 13
 / SOFTWARE: Patent in version 3.1
 / SEQ ID NO 9
 / LENGTH: 3094
 / TYPE: DNA
 / ORGANISM: Mus musculus
 / PUBLICATION INFORMATION:
 / DATABASE ACCESSION NUMBER: NM_008408
 / DATABASE ENTRY DATE: 2000-11-01
 / RELEVANT RESIDUES: (1)..(3094)
 / US-10-028-384-9

Alignment Scores:
 Pred. No.: 6,926-216 Length: 3094
 Score: 2198.50 Matches: 416
 Percent Similarity: 73.78% Conservative: 113
 Best Local Similarity: 58.02% Mismatches: 143
 Query Match: 54.34% Indels: 45
 DB: 15 Gaps: 8

US-10-028-384-8 (1-774) x US-10-028-384-9 (1-3094)

| | | | |
|----|-----|---|-----|
| QY | 17 | SerLeuIleThrPheAlaIleLeuLeuIleAlaIleLeuAlaGlyPheSerSerArgLeu | 36 |
| DB | 157 | ACACTCTAAAGCTCTCATCTCGATGCTGCTGTTATCTTTTCTACTCGTCT | 216 |
| QY | 37 | PheAlaValIleArgPheGluSerIleIleHicIleGluPheAspProThrPheAsnThrArg | 56 |
| DB | 217 | TTTGCTGTCGAGATTTGAAAGTGTCTCATCATGAGTTTGTATGCTACTTTAATATCGG | 276 |
| QY | 57 | AlaThrAlaIleMetValGlnAsnGlyIleThrPheAsnPheLeuAsnThrPheAspGluArg | 76 |
| DB | 277 | ACTACCCGGTTCTCGCTGAGAGGGGTTTATAAAATCCATAACTGTTTGTATGACCGG | 336 |
| QY | 77 | AlaThrProLeuGlyValGlyIleValGlyIleThrValIleProGlyLeuMetIleThr | 96 |
| DB | 337 | GCTTGTAACCTTTGGCCGGAATCATGGAGAACAAATTTACCAGGTTTAAATGACACT | 396 |
| QY | 97 | SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle | 116 |
| DB | 397 | TCTGTGCAATCTACCATGTACTCCATTTCTTCCATATCACTATTGACATTCGGAATGTC | 456 |
| QY | 117 | CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrIleLeuLeuThr | 136 |
| DB | 457 | TGTGTTTCCGGCCCACTTTTCTCTCTTCCACCACTCGTTACGTACCACTTACC | 516 |
| QY | 137 | LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaIleAspPheIleAlaIleValPro | 156 |
| DB | 517 | AAAGAGCTCAAGAGTGCAGAGAGCTGGGCTTCTGCTGCTGCATGATGCTGATGTTCT | 576 |
| QY | 157 | GlyIleIleSerArgSerValAlaGlySerIleAspAsnGluGlyIleAlaIlePheAla | 176 |
| DB | 577 | GGGTATATTTCTCGATCTGTAGCTGGCTCTTATGATATGAGGAATTTGATCTTTTGC | 636 |
| QY | 177 | LeuGlnPheThrIleThrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer | 196 |
| DB | 637 | ATGCTGCTTACTTACTATCATGTGATCAAGCACTGAAGCACTGGTTCCTCATATGGGCT | 696 |
| QY | 197 | AlaAlaAlaAlaLeuSerIlePheThrMetValSerAlaIleThrGlyIleValPheIle | 216 |
| DB | 697 | GCCAAAGTGGCCCTCGCTTATTTCTATCATGCTCTTCTTCTGAGGAGGCTATGTTCTCTG | 756 |
| QY | 217 | IleAsnLeuIleProLeuHisValPheValLeuLeuLeuMetGlyArgIleThrProArg | 236 |
| DB | 757 | ATCAACTTGATTTCTCTACATGCTGCTGGTGTAAATGCTGACAGGCCGTTTCTTCAACCG | 816 |
| QY | 237 | LeuLeuThrSerIleThrPheIleIleLeuGlyLeuLeuPheSerMetGlnIlePro | 256 |
| DB | 817 | ATCTACGTAGCTACTGTTACTGTTTACTGCTGGGACCACTTCTTCTTCTGAGATTC | 876 |


```

Qy 610 LeuValIlePheGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
Db 1846 CTTGTCTATTTTGGAGCCCTTACTGGGTATTTCTCGATGATGATCAACAAGTTCTTTGG 1905
Qy 630 MetValArgIle--AlaGluGlyGluHisProLysAspIleGlyGluSerAspTyrPhe 648
Db 1906 ATGCTCGGATTTGGAGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
Qy 649 ThrAspArgGlyGluPheArgValAlaGluGlyAlaProAlaLeuLeuAsnLysCysLeu 668
Db 1966 ACTCTACTGGGAATTCCTGTTGATCGCTGAGGGTCTCCGGTCTGCTCAACTGCTT 2025
Qy 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSer 688
Db 2026 ATGTACAAATGTACTACCGCTTTGGCAGGTCTACAGAGAGAGAGAGAGAGAGAGAG 2085
Qy 689 GlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeu 708
Db 2086 GCCTTTACCGCTGTTCGAAATGCTGAGATTGGTAAATGAGAGAGAGAGAGAGAGAGAG 2145
Qy 709 GluGluAlaTyrThrThrGluHisTyrLeuValArgIleTyrArgValLys 725
Db 2146 GAGGAGCGTATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2196

RESULT 6
US-10-417-375-95
; Sequence 95, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 52945201600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-95

Alignment Scores:
Pred. No.: 1,75e-215 Length: 3046
Score: 2194.50 Matches: 417
Percent Similarity: 73.47% Conservative: 112
Best Local Similarity: 57.92% Mismatches: 140
Query Match: 54.24% Indels: 51
DB: 18 Gaps: 9

US-10-028-384-8 (1-774) x US-10-417-375-95 (1-3046)

Qy 17 SerLeuIleThrPheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeu 36
Db 181 ACATTTTGAGCTTCTCATCTGTCAGAGGCTGCTGATATATCTTCTTCTTCTTCTGCTG 240
Qy 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTyrPheAsnTyrArg 56
Db 241 TTTGCTGCTCTGAGATTTGAAAGTGTATCCATGAGTTTGATCCCTGATCTTAAATATCGG 300
Qy 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuLeuAsnTyrPheAspGluArg 76
Db 301 ACTACCAAGTCTCTGCTGAGGAGGGGTTTATTAATTCATTAATCTGTTTGTATGACCGA 360
Qy 77 AlaTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
Db 361 GCCTGTACCTTTGGGAGAAATCATTTGGAGGAGAAATTTACCAAGGTTTAAATGATCACC 420
Qy 97 SerGlyGlyIleHisTyrPheLeuHisValLeuAsnIleProValHisIleArgAspIle 116
Db 421 TCTGCTGCAATCTACCATGCTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 480

```

```

Qy 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
Db 481 TGTGTGTCTCTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 540
Qy 137 LysGluLeuTyrSerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
Db 541 AAGAGAGCTCAAGGATGACAGGGCTGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Qy 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
Db 601 GGATATATCTCCGATCTCTGCTGGCTCTCTATGATGATGATGATGATGATGATGATGAT 660
Qy 177 LeuGlnPheThrTyrPheLeuTyrValArgSerValLysThrGlySerValPheTyrSer 196
Db 661 ATGCTACTACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 720
Qy 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrPheGlyTyrValPheIle 216
Db 721 GCTAAGTGTGGCTTGTCTTATTTCTATGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Qy 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
Db 781 ATCACTTAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
Qy 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
Db 841 ATCTATGTGGCTACTGCTACTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Qy 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
Db 901 TTTGTGGTGTTCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 961 CTCTGCCAGATCCATGCTCTTGTGATTAATCTCTGCGCAAGTGTGAATCCACCAACAT 1020
Qy 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
Db 1021 GAAGTCTCTTTCCGAGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1077
Qy 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTyrSerGlyArgPhe 331
Db 1078 -----GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1125
Qy 332 TyrSerLeuTyrPheThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
Db 1126 TACTGCTGCTGGATCCCTCTTATGCTAAGACACACATCCCATCTCTCTCTCTCTCTCT 1185
Qy 352 GluHisGlnProThrTyrPheSerPhePheAspLeuHisIleLeuValCysAla 371
Db 1186 GAGCATCAGCCCAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1245
Qy 372 PheProValGlyValTyrTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1246 TTTCCAGTGTGGCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1305
Qy 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1306 ATGTATGTTGTGACGAGCATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1365
Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1366 GCACCTGTATGTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1425
Qy 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
Db 1426 AAGAACTCTCGACATCAAGTCGTCCA-----GACAGAGAGAGCAAG----- 1449
Qy 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
Db 1450 -----GACAGAGAGAGCAAG-----GACAGAGAGAGCAAG----- 1464
Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489

```

Db 1465 -----AAGCACAGATCCACCTACCTATTAAGAATGAGTGGCAAGTGGG 1512
Qy 490 VALLLEuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
Db 1513 ATGATACTGGTCATGGCTTCTTCTCATCACTACACCTTCACTCAACCTGGGTGACC 1572
Qy 510 SerAsnAlaTyrSerProSerPheValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1573 AGTGAGGCTACTTCTCCGTCATCTGTCTATCTATCTGCTCCGCTGGTGGG---GATGGCAGT 1629
Qy 530 ArgAsnLeuLeuAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
Db 1630 AGGATCATATTGATGACTTCCGAGAGCAATATTATTGGCTTCGTATATATCTCCAGAG 1689
Qy 550 AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnLeuAlaGlyMetAlaAsnArg 569
Db 1690 GATCGGAAGGTCATGCTCGTGGTGGATATGGCTATCAGATTACAGTATGGCAACCGA 1749
Qy 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisLeuAlaLeuValGlyLysAla 589
Db 1750 ACAATTTTAGTGACAATAACACATGGAATAATACCATATTCTCGAGTAGGGCAGGCA 1809
Qy 590 MetSerSerThrGluGluLysSerTyrGluLeuMetThrSerLeuAspValAspTyrVal 609
Db 1810 ATGGCGTCCACAGAGGAAAGGCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGTG 1869
Qy 610 LeuValilePheGlyGlyValileGlyTyrSerGlyAspAspLeuAsnLysPheLeuTrp 629
Db 1870 CTGTCATTTTGGAGGCTCCTGCGTATCTCTGTGATATATCAACAGTTCTTTGG 1929
Qy 630 MetValArgileAla-----GluGlyGluHisProLysAspLeuLysGluSer 645
Db 1930 ATGGTCGGATTTGGAGGAGCACAGATACAGGCAACAT-----ATCAAGGAGAAT 1980
Qy 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
Db 1981 GACTATATATCTCACTGGGAGTCCGTTGGACCGTGAAGTTCTCCAGTGTGCTC 2040
Qy 666 AsnCysLeuMetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
Db 2041 AACTGCTCATGTACAAAGATGTCTACTATCGCTTTGGACAGTTTACACAGAGCCAG 2100
Qy 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValileGlyAsnLysAspPheAspLeu 705
Db 2101 CGTCCCTCCAGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTT 2160
Qy 706 ThrTyrLeuGluAlaTyrThrThrGluHisTrpLeuValArgileTyrArgValLys 725
Db 2161 GATGTCCTCGAGGAGCATATACACAGAACATTTGCTGGTCAGGATATACAGGTAAAG 2220

RESULT 7

US-10-417-375-99
; Sequence 99, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Maiandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-99

Alignment Scores: 5404
Pred. No.: 3,578-214 Length: 5404
Score: 2185.50 Matches: 415

Percent Similarity: 73.40% Conservative: 112
Best Local Similarity: 57.80% Mismatches: 140
Query Match: 54.02% Indels: 51
DB: 18 Gaps: 9

US-10-028-384-8 (1-774) x US-10-417-375-99 (1-5404)

Qy 17 SerLeuileThrPheAlaileLeuLeuileAlaTrpLeuAlaGlyPheSerSerArgLeu 36
Db 181 ACACITTTTCAAGCTTCTCATCTGTCATAGGTGCTGCTATTATCTCTCCACTCGTCTG 240
Qy 37 PheAlaValileArgPheGluSerileleHisGluPheAspProTrpPheAsnTyrArg 56
Db 241 TTGCTGTCTCAGAGATTGAAAGTGTATCCATGAGTTGATCCCTACTCTTTAAATATCGG 300
Qy 57 AlaThrAlaTyrMetValGlnAsnGlyTrpTyrAsnPheLeuAsnTrpPheAspGluArg 76
Db 301 ACTACCAGGTCTCTCGCTGAGGAGGGTTTATAAATCCATACTGGTTGATGACCGA 360
Qy 77 AlaTrpTyrProLeuGlyArgileValGlyGlyThrValTyrProGlyLeuMetileThr 96
Db 361 GCCTGGTACCCCTTTGGGACGAATCATTTGGAGAACAAATTTACCCAGGTTTAAATGATCACC 420
Qy 97 SerGlyGlyileHisTrpLeuLeuHisValLeuAsnileProValHisileArgAspIle 116
Db 421 TCTGCTGCATCTACCATGCTACTCTCATTTTTCACATCACCATCGACATTCGGATGTC 480
Qy 117 CysValPheLeuAlaProilePheSerGlyLeuThrSerileSerThrTyrLeuLeuThr 136
Db 481 TGTGTGTTCTTGGCCCT 540
Qy 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheileAlaileValPro 156
Db 541 AAAGAGCTCAAGGATCGAGGCGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 157 GlyTyrileSerArgSerValAlaGlySerTyrAspAsnGluGlyileAlailePheAla 176
Db 601 GGATATATCTCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Qy 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
Db 661 ATGCTACTCCTACTACTACTGATGATCAAGGCAAGAAAGACTGGTTCATCTGTGTGGCA 720
Qy 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIle 216
Db 721 GCTAAGTGTGCTTGTCTTCTATCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Qy 217 IleAsnLeuileProLeuHisValPheValLeuLeuileMetGlyArgTyrSerProArg 236
Db 781 ATCAACTTAATTCCTCTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Qy 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
Db 841 ATCTATGTGGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Qy 257 PheValGlyPheGlnProileArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
Db 901 TTTGTGGGTTTCCAGCCTGCTCTTTCATCAGAGCACATGCGAGCTTTGGGGTCTTTGGT 960
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 961 CTCTGCCAGATCCATGCTTTGTGGATTACCTGCGAGCAAGTGAATCCACACAAATTT 1020
Qy 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
Db 1021 GAAGTTCTTTCCGAGGCGTCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1077
Qy 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
Db 1078 -----GCTCTCTCATGCTGACA-----GGAAAAATATCTCCCTGGAGCGGCTTTC 1125
Qy 332 TyrSerLeuTrpAspThrGlyTyrAlaLysileHisileProileleAlaSerValSer 351

| | | | |
|----|------|--|------|
| Db | 1126 | TACTCGCTGCTGATCCCTCTTATGCTAGAACACACATCCCCATCATTCCTCTCTGCTCT | 1185 |
| QY | 352 | GluHisGlnProThrTrpPheSerPhePheAspLeuHisLeuValCysAla | 371 |
| Db | 1186 | GAGCATCAGCCCAACCTGGTCTCATACTATTTTGACTCGCAGCTCCTCGTCTTCATG | 1245 |
| QY | 372 | PheProValGlyValTrpTyrcysIleLysGlnIleAsnAspGluArgValPheValVal | 391 |
| Db | 1246 | TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTGTGATGCCGGATTTTATCATC | 1305 |
| QY | 392 | LeuTyrrAlaIleSerAlaValTyrrPheAlaGlyValMetValArgLeuMetLeuThrLeu | 411 |
| Db | 1306 | ATGTATGGTGTGACGACGATGACTTTTCAGCTGTAAATGGTGGCTGTAGTGTG | 1365 |
| QY | 412 | ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu | 431 |
| Db | 1366 | GCACTCTGTTATGTCATCTCTCTGTCANTTCCGAGTCTCCGAGTGTGTGCCACATCATG | 1425 |
| QY | 432 | GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu | 451 |
| Db | 1426 | AAGAACTGGACATAGTCTGCA----- | 1449 |
| QY | 452 | AlaGluAspSerIleGluLysLysThrLeuTyrrAspIleAlaGlyLysLeuLysHisArg | 471 |
| Db | 1450 | -----GACAAAGAGACGAA----- | 1464 |
| QY | 472 | ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle | 489 |
| Db | 1465 | -----AAGCAACAGGATTCACCTACCTATTAAAGTAAAGTGGCAATGGG | 1512 |
| QY | 490 | ValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThr | 509 |
| Db | 1513 | ATGATACTGGTCATGGCTTCTTCTTCATCACCTACACCTTTCATTCAACCTGGGTGACC | 1572 |
| QY | 510 | SerAsnAlaTyrrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer | 529 |
| Db | 1573 | AGTGAGGCCTACTCTTCCGTGCTCATGTGTAATCTCCCGTGGTGGG---GATGGCAGT | 1629 |
| QY | 530 | ArgAsnIleLeuAspAspPheArgGluAlaTyrrTyrrTripleuSerGlnAsnThrAlaAsp | 549 |
| Db | 1630 | AGGATCATATTGATGACTTCCGAGAGACATATTATTGGCTTCGTCAATAATCTCCAG | 1689 |
| QY | 550 | AspAlaArgValMetSerTrpTrpAspTyrrGlyTyrrGlnIleAlaGlyMetAlaAsnArg | 569 |
| Db | 1690 | GATCGGAAGGTCAATGCTCTGGTGGCATATTGGCTATCAGATTACAGTATGGCAACCGA | 1749 |
| QY | 570 | ThrThrLeuValAspAsnAsnThrTripleuAsnSerHisLeuAlaLeuValGlyLysAla | 589 |
| Db | 1750 | ACAATTTTAGTGGACATAACACATGGAATAATACCCATATTTCGAGTAGGGCAGGCA | 1809 |
| QY | 590 | MetSerSerThrGluGluLysSerTyrrGluIleMetThrSerLeuAspValAspTyrrVal | 609 |
| Db | 1810 | ATGGCGTCCACAGAGGAAAAGCCTATGAGATCATGAGGGAGCTCGATGTCAGCTATGTG | 1869 |
| QY | 610 | LeuValIlePheGlyGlyValIleGlyTyrrSerGlyAspAspIleAsnLysPheLeuTrp | 629 |
| Db | 1870 | CTGTGCTATTTTGGAGGCCTCACTGGGTATTTCCTCTGATGATATCAACAAGTTTCTTTGG | 1929 |
| QY | 630 | MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer | 645 |
| Db | 1930 | ATGTCCGGANTGGAGGAGGACACAGATACAGGCGCAACAT-----ATCAAGGACAAT | 1980 |
| QY | 646 | AspTyrrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu | 665 |
| Db | 1981 | GACTATTATATCCAACTGGGGAGTTCGGTGTGGACCGTGAAGTCTCCAGTCTGCTCTC | 2040 |
| QY | 666 | AsnCysLeuMetTyrrLysLeuSerTyrrTyrrArgPheGlyGluLeuLysLeuAspTyrrArg | 685 |
| Db | 2041 | AACTCCTCATGTACAGATATGTTATTCATTCCTTTGGACAGCTTTTACACAGAGCCCAAG | 2100 |
| QY | 686 | GlyProSerGlyTyrrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu | 705 |
| Db | 2101 | CGTCTCCAGGCTTTGACCGTGTCCGAATGCTGAGATTGGGAATAAGAATTTGAGCTT | 2160 |

Qy 706 ThrTyrLeuGluAlaTyrThrThrGluHisTrpIeuValArgIleTyrArg 723
| | | | |
Dh 2161 GATGTCTGGAGGAAGCATATACCAAGAACATTGGCTGGTCAGGATATACAAG 2214

RESULT 8

```

US-10-417-375-97
; Sequence, 97, Application US/10417375
; Publication NO. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 5827
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-97

```

| | | | |
|------------------------|---------|---------------|------|
| Alignment Scores: | 4e-214 | Length: | 5927 |
| Pred. No.: | 2185.50 | Matches: | 415 |
| Score: | 73.40% | Conservative: | 112 |
| Percent Similarity: | 73.40% | Mismatches: | 140 |
| Best Local Similarity: | 57.80% | Indels: | 51 |
| Query Match: | 54.0% | Gaps: | 9 |
| DB: | 1A | | |

115-10-038-384-8 (1-774) x 115-10-417-375-97 (1-5827)

| | | | | |
|--|----|-----|--|-----|
| | Qy | 17 | SerLeuIleThr-PheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerArgIeu | 36 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 181 | ACACTTTGAAGCTTCTCATTCGTCAATGGCTGCTATTATCCTTCACCACCTGCTG | 240 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 37 | PheAlaValIleArgPheGluseriIleHisclupheAspProTrpPheAsnTyArg | 56 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 241 | TTTGCTGTCCTGAGATTGAAAGTGTTATCATCAGATTGCATCGTACTTTAATTATCGG | 300 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 57 | AlaThrAlaTyrrMetValGlnAsnClYrTrpTyrrAsnPheLeuAsnTrpPheAspGluArg | 76 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 301 | ACTACCAGGTCTCTGGCTGAGGAGGGGTTTTATAATTCATTAACCTGGTTTGAATGCCGA | 360 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 77 | AlaTrpTyrr-ProLeuGlyArgIleValGlyGlyThrValTyrrProGlyLeuMettileThr | 96 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 361 | GCTGTGTAACCCITGGRCAGATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCAC | 420 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 97 | SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle | 116 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 421 | TCTGCTGCAATCTACCATGTATCCTCATTTTTTCCACATCCACATCGCATTCGGAATGC | 480 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 117 | CysValPheIleuAlaProIlePheSerGlyLeuThrSeriIeSerThrTyrrLeuLeuThr | 136 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 481 | TGTGTGTTCCTGGCCCCCTCTCTCTCCCTTCCACCACCATCGTCAGTACCACTTACC | 540 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 137 | LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro | 156 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 541 | AAGAAGCTCAAGGATCGAGGGCTGGGCTTCTTGCTGTGCCATGATGTGCTAGATTCTCT | 600 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 157 | GlyTyrrIleSerArgSerValAlaGlySerTyrrAspAsnGluGlyIleAlaIlePheAla | 176 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 601 | GCATATATCTCCCGATCTGTGGCTGGCTCCTATGATATGAAGGGATGTCATCTTTTGC | 660 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 177 | LeuGlnPheThrTyrrPheLeuTrpValArgSerValIlysThrGlySerValPheTrpSer | 196 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 661 | ATGTACTACCTACTACATGTGATCAAGCGATPAAGACTGGTTCCATCTGTTGGCCA | 720 |
| | | | ::: ::: ::: ::: ::: | |
| | Qy | 197 | AlaAlaAlaAlaLeuSerTyrrPheTyrrMetValSerAlaIatPrpGlyTyrrValPheIle | 216 |
| | | | ::: ::: ::: ::: ::: | |
| | Dd | 721 | GCTAACGTCGCTTCTGCTTATTTTACATGGTCTCGTCACTGGGAGGATATGTGTTCTGT | 780 |
| | | | ::: ::: ::: ::: ::: | |

| | | | |
|----|------|--|------|
| Qy | 217 | IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg | 236 |
| Db | 781 | ATCAACTTAATTCCTCCACGTCCTCGTGTGATGCTCAGAGCGGTTTCTTCACCGG | 840 |
| Qy | 237 | LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro | 256 |
| Db | 841 | ATCTATGTGGGCTACTGTACTGTTTACTGCGCTGGCACATATACTTCTCATGCAGATCTCC | 900 |
| Qy | 257 | PheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeuGlyValPheVal | 276 |
| Db | 901 | TTTGTGGGTTCCAGCGTCTCTTCATCAGAGCATGGCAGCCTTTGGGGTCTTTGGT | 960 |
| Qy | 277 | LeuLeuMetAlaValAlaThrLeuAArgHisLeuGlnSerValLeuSerArgAsnGluPhe | 296 |
| Db | 961 | CTCTGCCAGATCCATGCCTTTGTGTGATTAACCTGGCAGCAAGTTGAATCCACAACAATTT | 1020 |
| Qy | 297 | ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal | 311 |
| Db | 1021 | GAAGTTCTTTTCGGAGGCTCATCTCTGTGTAGCCTTGTCCTTCACCGTGGGA--- | 1077 |
| Qy | 312 | PheValAlaValValLeuThrMetLeuGlyValValAlaLapProTyrSerGlyArgPhe | 331 |
| Db | 1078 | -----GCTCTCTCATGCTGACA-----GGAATAATATCTCCTCGGACGGGCGTTTC | 1125 |
| Qy | 332 | TyrSerLeuTrpAspThrGlyTyrAlaIleHisIleProIleIleAlaSerValSer | 351 |
| Db | 1126 | TACTCGTGTGGATCCCTCTTATGCTAAGAACCAACATCCCATCATGTGTTCTGTCT | 1185 |
| Qy | 352 | GluHisGlnProThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla | 371 |
| Db | 1186 | GAGCATCAGCCCAACCTGGTCTCATACTATTATTGACCTGCGAGCTCTCGTCTTCATG | 1245 |
| Qy | 372 | PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal | 391 |
| Db | 1246 | TTTCCAGTTGGCCTCTAATCTACTGCTTAPGCAACCTGCTCATGCCCCGAGATTTCATCATC | 1305 |
| Qy | 392 | LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu | 411 |
| Db | 1306 | ATGTATGGTGTGACCAGCATGACTTTTCAGCTGTAAATGGTGGTCTTAATGCTAGTGTG | 1365 |
| Qy | 412 | ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu | 431 |
| Db | 1366 | GCACCTGTTATGTGCATCTCTCTGGCATTTGGAGTCTCCAGGTGCTGTCCACATACATG | 1425 |
| Qy | 432 | GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu | 451 |
| Db | 1426 | AAGAACTGGACATAAGTCGTCCA----- | 1449 |
| Qy | 452 | AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg | 471 |
| Db | 1450 | -----GACAAGAAGACGAAG----- | 1464 |
| Qy | 472 | ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle | 489 |
| Db | 1465 | -----AAGCAACAGGATTCACCTACCTTAAAGTAAGTAAGTGCGCACTGGG | 1512 |
| Qy | 490 | ValIleLeuAlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThr | 509 |
| Db | 1513 | ATGATACTGTCTGCTGCTTCTTCTCATCACCTACACCTTTTCATTCACTTGGTGACC | 1572 |
| Qy | 510 | SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer | 529 |
| Db | 1573 | AGTGAGCGCTTACTTCTTCGTCCATGTACTATCTGCGCGTGGTGGG---GATGCGAGT | 1629 |
| Qy | 530 | ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTripleSerGlnAsnThrAlaAsp | 549 |
| Db | 1630 | AGGATCATATTTGATGACTTCCGAGAAGCATATTAATGGCTTCGTATATACTCCAGAG | 1689 |
| Qy | 550 | AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg | 569 |
| Db | 1690 | GATGCCAAGGTCAATGCTTGGTGGGATATGGCTATCAGATTACAGTATGGCAACCGA | 1749 |

| | | | |
|----|------|--|------|
| QY | 570 | ThrThrLeuValAspAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla | 589 |
| Db | 1750 | ACAATTTTAGTGGACAATAACACATAGGAATAATACCCATATTTCTCCAGTAGGCAGGCA | 1809 |
| QY | 590 | MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal | 609 |
| Db | 1810 | ATGGCGGTCCACAGAGAAAAGCCTATGAGATCATGAGGAGCTCGGATGTGAGCTATGTG | 1869 |
| QY | 610 | LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp | 629 |
| Db | 1870 | CTGGTCATTTTGGAGGCTCACTGGGTATTCTCTGATGATATCAACAAGTTCTTTGG | 1929 |
| QY | 630 | MetValArgIleAlaA-----GluGlyGluHisProLysAspIleLysGluSer | 645 |
| Db | 1930 | ATGGTCCCGATGGAGGAGCACAGATACAGGCAACAT-----ATCAAGAGCAAT | 1980 |
| QY | 646 | AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu | 665 |
| Db | 1981 | GACTATTATATCTCCAACCTGGGAGTTCGGTGTGGACCGTGAAGTTCTCAGTGTCTGCTC | 2040 |
| QY | 666 | AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg | 685 |
| Db | 2041 | AACTGCCTCATGTACAAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAAGCCCAAG | 2100 |
| QY | 686 | GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu | 705 |
| Db | 2101 | CGTCTCCAGCCTTTTGACCGGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTT | 2160 |
| QY | 706 | ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArg | 723 |
| Db | 2161 | GATGTCCTGGAGGAAGCATATACACAGACATTTGCTGCTCAGGATATACAAG | 2214 |

RESULT 9

US-10-171-581-112

Sequence 112, Application US/10171581

Publication No. US20030104426A1

GENERAL INFORMATION:

APPLICANT: Dai, Hongyue

APPLICANT: Linsley, Peter

APPLICANT: Mac Mao

TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia

FILE REFERENCE: 9301-157-999

CURRENT APPLICATION NUMBER: US/10/171,581

CURRENT FILING DATE: 2002-06-14

PRIOR APPLICATION NUMBER: 60/298,914

PRIOR FILING DATE: 2001-06-18

NUMBER OF SEQ ID NOS: 366

SEQ ID NO 112

LENGTH: 2472

TYPE: DNA

ORGANISM: Homo sapiens

PUBLICATION INFORMATION:

DATABASE ACCESSION NUMBER: L38961

DATABASE ENTRY DATE: 2001-06-18

US-10-171-581-112

Alignment Scores:

Pred. No.: 7,22e-214 Length: 2472

Score: 2177.50 Matches: 415

Percent Similarity: 72.92% Conservative: 110

Best Local Similarity: 57.64% Mismatches: 144

Query Match: 53.82% Indels: 51

DB: 15 Gaps: 9

US-10-028-384-8 (1-774) x US-10-171-581-112 (1-2472)

US-10-028-384-8 (1-774) x US-10-171-581-112 (1-2472)

| | | | | |
|----|--|-----|---|-----|
| Cy | | 17 | SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeu | 36 |
| Dd | | 152 | ACAATTTCGAAGCTTCTCATCTGTCAATGGTGCTGTTATTCCTTCACACGCCTG | 211 |
| Cy | | 37 | PheAlaValIleArgPheCysIleSerIleIleHisGluPheAspProTrpPheAsnTrpArg | 56 |
| Dd | | 212 | TTCGCGTCCCTCAGATTTGAAAGTCTTATCATGTAGTTTGTACGTACTTTAATTATCGG | 271 |

FILE REFERENCE: 5600-74
 CURRENT APPLICATION NUMBER: US/10/028.384
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 13
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 11
 LENGTH: 2472
 TYPE: DNA
 ORGANISM: Homo sapiens
 PUBLICATION INFORMATION:
 DATABASE ACCESSION NUMBER: NM_002219
 DATABASE ENTRY DATE: 2000-10-31
 RELEVANT RESIDUES: (1)..(2472)
 US-10-028-384-11

Alignment Scores:
 Pred. No.: 7,22e-214 Length: 2472
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 53.82% Indels: 51
 DB: 15 Gaps: 9

US-10-028-384-8 (1-774) x US-10-028-384-11 (1-2472)

```

QY 17 SerLeuLeuThrPheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerSerArgLeu 36
Db 152 ACATTTTGAAGCTTCTCATCTGCTCAATGGCTGCTGATATTATCTCTCCACTCGTCTG 211
QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 56
Db 212 TTGCTCTCTGAGATTGAAAGTGTATCCATGAGTTGATCCGACTTAAATATCGG 271
QY 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTrpTyrAsnPheLeuAsnTrpPheAspGluArg 76
Db 272 ACTACCAAGTTCTCGTGGCTGAGGAGGGGTTTATAAATTCATTAACCTGTTGATGACCGA 331
QY 77 AlaTrpTyrProLeuGlyArgIleValGlyThrValTyrProGlyLeuMetIleThr 96
Db 332 GCCTGTGTACCTTTGGACCAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACC 391
QY 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
Db 392 TCTGTGCAATCTACCATGTAACCTCAATTTTCCACATCCACATCGACATTCGGAATGTC 451
QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
Db 452 TGTGTGTCTGCGCCCTCTCTTCTCCTCCTCCACCTCCATCGTACAGTACCTCTCTTACC 511
QY 137 LysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
Db 512 AAAGAGCTCAAGGATGCAGGGGCTGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 571
QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
Db 572 GGAATATATCTCCGATCTGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631
QY 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValIleThrGlySerValPheTrpSer 196
Db 632 ATGCTACTACCTACTACATGATGGATGAGGAGGATGAGGAGGATGAGGAGGATGAGGAGG 691
QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
Db 692 GCTAAGTGTGCCCTTCTATTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
Db 752 ATCACTTAATTCCTCTCACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811
QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
Db 812 ATCTATGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 871

```

```

QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
Db 872 TTGTGGGTTTCCAGCCTGCTCTTTCATCAGACACATGCGAGGCTTTGGGCTCTTGGT 931
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 932 CTCTGCCAGATCCATGCCCTTTGTGGATACCTGCCAGCAAGTGAATCAACCAACATTT 991
QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
Db 992 GAAGTTCTTTTCCGGAGCGCTCATCTCTGTGGTAGGCTTTGTCTCTCTCACCGTGGGA--- 1048
QY 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
Db 1049 -----GCTCTCTCATGCTGACA-----GAAAAAATATCTCCCTGACCGGGGCTTTC 1096
QY 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
Db 1097 TACTCACTGCTGATCCCTCTTATGCTTAAGAACACATCCCATCATTTCTCTCTGCTCT 1156
QY 352 GluHisGlnProThrThrTrpPheSerPhePhePheAspLeuHisIleLeuValCysAla 371
Db 1157 GAGCATCAGCCCAACACCTGCTCTCATATTTTGACCTTGACCTCTCTCTCTCTCATG 1216
QY 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1217 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCCCGGATTTTATCATC 1276
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1277 ATGTATGTGTGACCAAGCATGCTATTTTCAGCTGTATGTCGCTCTAATGCTAGTGTG 1336
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1337 GCACCTGTTATGAGCATTTCTCTGCGCATTCGAGTCTCCAGGCTGCTCCACATACATG 1396
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaIleThrGluValAspGlu 451
Db 1397 AAGAATCTGGACATAAAGTCGCCCA----- 1420
QY 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
Db 1421 -----GACAAAGAGACAAG----- 1435
QY 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
Db 1436 -----AAGCAACAGGATTCACCTTATTAAGATTGAAGTGGCAAGTGGG 1493
QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
Db 1494 ATGATCTGGTCATGGGCTTTCTTCTCATCACCTACACCTTTTCATTCACCTGGGTGACC 1543
QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1544 AGTGAGGCTACTCTTCTCGTCCATTGTACTATCTCCCGTGGTGGG---GATGGCAGT 1600
QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
Db 1601 AGGATCATATTGATGACTTCCGAGAAAGCATATTATTGGCTTGTCTCATATACTCCAGAG 1660
QY 550 AspAlaArgValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
Db 1661 GATCGAAGGTCATGCTCTGCTGGGATTTGGCTATCAGATTACAGCTATGGGAAACCGA 1720
QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
Db 1721 ACAATTTTAGTGGACAATAACATGGAAATAATACCATATTTCCTCGAGTAGGGCAGCA 1780
QY 590 MetSerSerThrGluGlyLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
Db 1781 ATGCGGTCACAGAGGAAAGCCTATGAGATCATGAGGAGGAGTGTGATGTCACCTATGTG 1840
QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp 629

```


Db 1841 CTGGTCATTTTGGAGGCTCACTGGGATTCCTCTGATGATATCAACAAGTTTCTTTGG 1900
Qy 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
Db 1901 ATGTCCTGGAGTGGAGGACACAGATACAGGCAACAT-----ATCAAGGAGAT 1951
Qy 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
Db 1952 GACTATTATCTCAACTCACTGGAGTTCCGTGTGACCGTGAAGTTCTCCAGTCTCTC 2011
Qy 666 AsnCysLeuMetTyrLysLeuSerTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
Db 2012 AACTGCCCATGTACAGATGTGTACTATCGTTTGGACAGGTTTACACAGAGCCCAAG 2071
Qy 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
Db 2072 CGTCTCCAGGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTT 2131
Qy 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
Db 2132 GATGCTCGGAGGAGGCTATACACAGAACATTCGCTGGTCAGGATATACAGGTAAG 2191

RESULT 11

US-10-172-118-742
; Sequence 742, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002219
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-742

Alignment Scores:

Pred. No.: 7,226-214 Length: 2472
Score: 2177.50 Matches: 415
Percent Similarity: 72.92% Conservative: 110
Best Local Similarity: 57.64% Mismatches: 144
Query Match: 51 Indels: 51
DB: 15 Gaps: 9

US-10-028-384-8 (1-774) x US-10-172-118-742 (1-2472)

Qy 17 SerLeuIleThrPheAlaIleLeuLeuIleAlaTrpLeuAlaGlyPheSerArgLeu 36
Db 152 ACATTTTGAAGTTCTCATCTGTCAATGGCTGCTGTATATCTTCTCCACTGCTG 211
Qy 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProTrpPheAsnTyrArg 56
Db 212 TTTCGTCTCTGAGATTGAAAGTGTATCCATGATTTGATCGGTACTTTAATATCG 271
Qy 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArg 76
Db 272 ACTACCAGGTTCTGCTGGCTGAGGAGGGGTTTATAAATTCATAACTGGTTTGTACCGA 331

Qy 77 AlaTrpTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
Db 332 GCCTGGTACCCCTTTGGAGCAATCATTTGAGAACATTTACCCAGGTTTAAATCATCAC 391
Qy 97 SerGlyGlyIleHisThrLeuHisValLeuAsnIleProValHisIleArgAspIle 116
Db 392 TCTGTGCAATCTACCATGTACTCTCATTTTCCACATCACCATCGACATTCGGAATGTC 451
Qy 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
Db 452 TGTGTGTTCTGGCCCTCTCTTCTCCCTTCCATCCATCGTCAGTACTCTCTTACC 511
Qy 137 LysGluLeuTrpSerAlaGlyValaGlyLeuPheAlaIleAspPheIleAlaIleValPro 156
Db 512 AAAGAGCTCAAGGATGAGGGCTGGCTTCTTGTCTGCTGCATGATGCTGTAGTTCTCT 571
Qy 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
Db 572 GGATATATCTCCGATCTGTGGCTGGCTCCTATGATGATGAAGGATTCGCATCTTTTGC 631
Qy 177 LeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
Db 632 ATGCTACTCACCTACTACATGTGGATCAAGGAGTAAGAGCTGTTCCATCTGTTGGCA 691
Qy 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGlyTyrValPheIle 216
Db 692 GCTAAGTGTGGCTTGTCTTATTTCTACATGTTCTGCTCATGGGAGGTTATGTGTTCTG 751
Qy 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArg 236
Db 752 ATCAACTTAAATTCCTCTCCAGCTCTCTGCTGTGATGCTCACAGGCGGTTTCTCACCG 811
Qy 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
Db 812 ATCTATGTGGCTACTGTTACTGCTGGGTACTATATCTTCTAGGCAGATCTCC 871
Qy 257 PheValGlyPheGluProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
Db 872 TTTGTGGGTTTCCAGCTCTCTCTTTCATCAGAGCACATGGCAGGTTTGGGCTCTTGGT 931
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 932 CTCTGCCAGATCCATGCTTTGTGGATTACCTGGCAGCAAGTTGAATCCACAATTT 991
Qy 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
Db 992 GAAGTTCTTTTCCGAGCGTCATCTCTGTTAGGCTTTGTCTCTTCTACCCGTGGGA--- 1048
Qy 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
Db 1049 -----GCTCTCTCATGCTGACA-----GGAAAAATATCTCCCTGGACGGGGTTTC 1096
Qy 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAspValSer 351
Db 1097 TACTCACTGTGGATCCCTCTTATGTAAAGAACACATCCCATCATCTCTCTGTGTCT 1156
Qy 352 GluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
Db 1157 GAGCATCAGCCCAACCTGCTCTCTATCTATTGACTGCTGAGCTCTCTCTCTCANG 1216
Qy 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1217 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTCTCTGATGCCCGGATTTTATCATC 1276
Qy 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1277 ATGTATGTTGTCAGGACGATGTTCTTTCAGCTGTATGTGGCTCTATGCTAGTGTG 1336
Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1337 GCACCTGTATGAGCAATCTCTCTGGCATTTGGAGTCTCCAGGCTGCTGCCACATACATG 1396

Db 812 ATCTATGTCGCTACTGTACTGTTTACTGCTGGTACTATATCTTCTAGGACGATCTCC 871
Qy 257 PheValGlyPheGlnProIleAthrSerGluHisMetAlaAlaLeuGlyValPheVal 276
Db 872 TTTGTGGTTCAGGCTGCTCTTCTATCAGACGACATGGCAGGGTTCGCTTGGT 931
Qy 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
Db 932 CTCTGCCAGATCCATGCTTTGTGGATTACCTGCGCAGCAAGTTGAATCCCAACAATTT 991
Qy 297 ArgIysLeuPhe-----IleValGlyGlyLeuValGlyValGlyVal 311
Db 992 GAAGTTCTTTTCGGAGCGTCATCTCTGTTAGGCTTTGCTCTCTCACCGTGGGA-- 1048
Qy 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTyrSerGlyArgPhe 331
Db 1049 -----GCTCTCTCATGCTGACA-----GGAAAAATATCTCCCTGGACGGGCGTTC 1096
Qy 332 TyrSerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
Db 1097 TACTCACTGCTGATCCCTCTTATGCTAAGAACACATCCCATCATTCCTCTGTCCT 1156
Qy 352 GluHisGlnProThrThrPheSerPhePhePheAspLeuHisIleLeuValCysAla 371
Db 1157 GAGCATCAGCCCAACACCTGCTCTCATATCTTTTACCTTCGACCTCTCTCTTCATG 1216
Qy 372 PheProValGlyValTyrTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
Db 1217 TTTCCAGTGGCTCTATCTATCTGTTAGCAACCTGCTGATGCGCGGATTTTATCATC 1276
Qy 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
Db 1277 ATGTATGCTGTGACGATGCTCTGCGATGCGAGTGGAGTCTCCAGTCTCTCATGCTGTG 1336
Qy 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
Db 1337 GCACCTGTTATGACATCTCTGCGATGCGAGTGGAGTCTCCAGTCTCTCCACATACATG 1396
Qy 432 GlnGluAspSerSerIysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
Db 1397 AAGAACTGACATAAGTCGCCCA----- 1420
Qy 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
Db 1421 -----GACAGAGAGCAAG----- 1435
Qy 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
Db 1436 -----AAGCAACAGGATTCACCTACCTTATTAAGATTGAAGTGGCAAGTGGG 1483
Qy 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTyrValThr 509
Db 1484 ATGATACGTGTCATGCTTCTTCTCATCACCTACCTTCTCATCAACCTGGGTGACC 1543
Qy 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
Db 1544 AGTGAGCCCTACCTCTTCCTGCTCAATGTACTATCTGCCGCTGGTGGG-----GATGGCAGT 1600
Qy 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTyrTyrLeuSerGlnAsnThrAlaAsp 549
Db 1601 AGGATCATATTCATGACTTCGAGAGCATATATTGCTTCGTCATAAATCTCCAGAG 1660
Qy 550 AspAlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
Db 1661 GATGCGAAGTCAATGCTGCTGGGATATGCTATCATGATATACAGTATTCGCAACCGCA 1720
Qy 570 ThrThrLeuValAspAsnAsnThrTyrAsnAsnSerHisIleAlaValGlyLysAla 589
Db 1721 ACAATTTTACTGACCAATAACACATGCAATAATACCATATTTCTCGAGTAGGCGACGA 1780
Qy 590 MetSerSerThrGlnGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
Db 1781 ATGCGCTCCACAGAGAAAGCCATGATGATCATGAGGAGCTCGATGTCAGCTATGTG 1840

Qy 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyr 629
Db 1841 CTGGTCATTTTGGAGGCTCTCACTGGGTATTTCTCTGATATCAACAAGTTCTTTGG 1900
Qy 630 MetValArgIleAla-----GluGlyGluHisProLysAspIleLysGluSer 645
Db 1901 ATGGTCCCGATTTGGAGGAGGACAGATACAGGCAACAT-----ATCAAGGAGAAT 1951
Qy 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
Db 1952 GACTATTATCTCCAACTGGGGAGTTCGCTGGACCGTGAAGTTCCTCCAGCTGCTC 2011
Qy 666 AnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
Db 2012 AACTGCTCATGTACAAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGCAAG 2071
Qy 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
Db 2072 CGTCTCCAGGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTT 2131
Qy 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisThrLeuValArgIleTyrArgValLys 725
Db 2132 GATGCTCGAGGAAGGCTATACACAGAACTTGGCTGCTCAGGATATACAGGTAAAG 2191
RESULT 13
US-10-032-585-6323
; Sequence 6323, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 6323
; LENGTH: 2256
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6323
Alignment Scores:
Pred. No.: 8,96e-214 Length: 2256
Score: 2176.00 Matches: 434
Percent Similarity: 70.16% Conservative: 109
Best Local Similarity: 56.07% Mismatches: 169
Query Match: 53.78% Indels: 62
DB: 15 Gaps: 15
US-10-028-384-8 (1-774) x US-10-032-585-6323 (1-2256)
Qy 11 LysValAlaGlyTyrSer---SerLeuIleThrPheAlaIleLeuLeuIleAlaTyrLeu 29
Db 34 AAAGTATTGGGATAGATGTTGAACCTGATTAGAGTATTTTAAAGTATTTATTTATA 93
Qy 30 -----AlaGlyPheSerArgLeuPheAlaValIleArgPheGluSerIle 45
Db 94 TCATAGACAGGTGCGCTATTTCTTCGTTTATTTTCGCTGATTCGATTTGAAGTATT 153
Qy 46 IleHisGluPheAspProTyrPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGly 65
Db 154 ATTATGAATTCGATCTCTGTTCAATTCGAGCAACCAATATTTAGTCACTCATTC 213
Qy 66 TPTTyrAsnPheLeuAsnThrPheAspGluArgAlaTyrTyrProLeuGlyArgIleVal 85
Db 214 TTTTGAATTTTGAATTTGTTGATGATAGAACTTGGTACCATTTGGGAAGTCACT 273
Qy 86 GlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTyrLeuLeuHis 105

Db 274 GGTGGTACTTTATATCCGGTTTAAATGGTGAATTCAGGTGCCATT--TGG-----CAT 324
 Qy 106 ValLeuAsn-----IleProValHisIleArgSpileCysValPheLeuAla 121
 Db 325 ATTTTACGTGATTGGTTGGCTTACCCGTTGATATTAGAAATATTGTGTATTATTAGCA 384
 Qy 122 ProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTyrSer 141
 Db 385 CCAGTTTCTCGGGATTAACTGCAATTGTGACTATTATTTTGTACTAAAGAAATGAAGGAT 444
 Qy 142 AlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArg 161
 Db 445 TCTAGTCAGAGATTATTGGCAGCTATATTATTCGGGATTGCCAGGTTATATTTCAGA 504
 Qy 162 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyr 181
 Db 505 TCAGTGGCTGGTCTTATGATGATGAAGCAATTGCCATTACTATTATTATTAATGGCAATTT 564
 Qy 182 PheLeuTyrValArgSerValLysThrGlySerValPheThrPheSerAlaAlaLeu 201
 Db 565 TATTCTCGGAATTAATCAATGAATAATGGGTTCAGTTTCTATGCCACATTGCAGCATTA 624
 Qy 202 SerTyrPheTyrMetValSerAlaTyrGlyTyrValPheIleIleLeuLeuIlePro 221
 Db 625 TTCTATTCTATATGTTAGTGGCTTGGGTGGATATGTTTCAATACCAATTTGATTCGA 684
 Qy 222 LeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSerTyr 241
 Db 685 TTACACGTAATTTGCTTGTATTTTTCATGGGTGCTTATAATGCCAAACTTTACACTGCTTAT 744
 Qy 242 SerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGln 261
 Db 745 ACTACATGTTGGCTTGGTACTTTGGCATCAATGCAGATCCCATTCGTTGGGTTTTTA 804
 Qy 262 ProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaVal 281
 Db 805 CCAATTAGATCAATCATCATATGGCTGCATTAGGAGTATTGGATTGTGTACAAATTAGTG 864
 Qy 282 AlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPheIle 301
 Db 865 GCTTTTGGTGAATTTATGTTAAATCAAAAGTTCACCAACAAATTTAAATTCATCTTGATA 924
 Qy 302 ValGly---GlyLeuLeuValGlyValGlyValPheValAlaValValLeuThrMet 320
 Db 925 GTTTCATGTTACTTGTGTGGATTAGGTATT--GGTGGATTATTGGATTACAGCA 981
 Qy 321 LeuGlyValValAlaProThrSerGlyArgPheTyrSerLeuThrAspThrGlyTyrAla 340
 Db 982 ATGGGTGGATTGCTCTTGGACAGGTAGATTATTATTCCTTATGGATACAAATTTATGCC 1041
 Qy 341 LysIleHisIleProIleAlaSerValSerGluHisGlnProThrThrTyrPheSer 360
 Db 1042 AGAATTCATATTCCAATATTCCTTCTGTTCTGAACTCAACCTACTGCTTGGCCAGCA 1101
 Qy 361 PhePhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrPheCysIle 380
 Db 1102 TTCCTTTTCGATACTAGTATGCTTATTGGTTATTCCCGCTGGTATCTATTATTGTTTC 1161
 Qy 381 LysGlnIleAsnAspGluArgValPheValValLeuThrAlaIleSerAlaValTyrPhe 400
 Db 1162 CAAGAATTGAAGGATGAACACGTTTTCATTTATCATTCAGTGTATTGTGTTCTTATTATT 1221
 Qy 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
 Db 1222 GCTGGTGTATGCTTAAGATTGATGTTGACTTTGACTTCCAGTCATTTGTTGCTGCGAGCA 1281
 Qy 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgValGly 440
 Db 1282 ATTGCTTATCTAAATTTGTTGATGCTATTG-----1314
 Qy 441 ThrAlaIleSerAlaAlaThrGluValAlaAspGluAlaGluAspSerIleGluLysThr 460

Db 1315 -----GACATTGTTGATTGTTCTACT 1335
 Qy 461 LeuTyrAspLysAlaGlyLysLeuLysHisIleThrLysHisAspAlaGlnIleAspThr 480
 Db 1336 -----GAGAAAGTTGAAAGTATGATGATGACGTTAGTACGAATCCAAGAGCTCAACC 1389
 Qy 481 GlyValSerSerAsn-----LeuLysSerIleValIleLeu 492
 Db 1390 AAAAAATCAAGTTCCAGATTTCGAATTCCTGATATTTGTCAAAAGTTTGTGTTTACTG 1449
 Qy 493 AlaValLeuMetLeuLeuMetMetPheAlaValHisCysThrTyrValThrSerAsnAla 512
 Db 1450 ACAATTTACATTTTACCTTTTCTACTTTTACATGTTTACATGTTGGTAAACATCGATGCT 1509
 Qy 513 TyrSerSerProSerIleValLeuAlaPheHisIleSerGlnAspGlySerArgAsnIle 532
 Db 1510 TATTTCATCACCATTGATTTTGTAGCTCCAGAAC---CCAGATGGCTCACAACATATC 1566
 Qy 533 LeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAspAspAlaArg 552
 Db 1567 ATTGATGATTATAGAGAGCTTATTCTGTTTGAATGAATACACAGAGATGCCAAA 1626
 Qy 553 ValMetSerTyrTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrThrLeu 572
 Db 1627 GTTATGGCTGGTGGGATTATGTTTATCAATCCGGGGTATGGCTGATAGAACACACTT 1686
 Qy 573 ValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
 Db 1687 GTTGATAACATACATGGAATACACACATATTCCTACTGTTGGTAAGGCAATGCTCTCC 1746
 Qy 593 ThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrValLeuValIle 612
 Db 1747 CCTGAAGATGTCGTATGAAATTTTGAGACACACACGATGTTGATTATGTAGTTATA 1806
 Qy 613 PheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArg 632
 Db 1807 TTTGGAGGTTATTGGGTTTATTCTGGTGATGATTAACAAATTCCTTATGGATGGTAAGA 1866
 Qy 633 IleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThrAspArgGly 652
 Db 1867 ATTGCTCAAGTATCTGGCTGTGTAATCAAGAAAGAGACTACTTTACTGACCCGAGGA 1926
 Qy 653 GluPheArgValAlaAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMetTyrLysLeu 672
 Db 1927 GAATPATAAGTGGATAAAGATGCATCTCGCAATGAAGAAATTCCTTGTATGATAAGTTA 1986
 Qy 673 SerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGlyTyrAspArg 692
 Db 1987 TCGTATCATAGATTCACTGATTTG-----TTTGGAGTAGAGATGTTGTTGATAGA 2037
 Qy 693 ThrArgAsnAlaValIle---GlyAsnLysAspPheAspLeuThrTyrLeuGluGluAla 711
 Db 2038 GTTAGAAACCAACAAATCCAGCCAATGAAGTACCGAAATTTGAATGTTGTGAAGAAGCC 2097
 Qy 712 TyrThrThrGluHisTyrPheValValArgIleTyrArgValLysLysProHisGluPheAsn 731
 Db 2098 TTCATCATCAAAATTTGGATTGAGATTTACAAAGTTTAA-----2139
 Qy 732 ArgProSerLeuLysThrLysGluArgThrIleProProAlaAsnPheIleSerArgLys 751
 Db 2140 -----GATTGGATAATGTTGGTAGAGATTTACATCAAGCTACTGCTTTTGAAGAATCA 2193
 Qy 752 AsnSerLysArgArgLysGlyTyrIleArgAsnArgProVal 765
 Db 2194 TCATCCGCACTTCCAAA-----AGAAACAGATCCATA 2226

RESULT 14
 US-10-128-714-7139
 ; Sequence 7139, Application US/10128714
 ; Publication No. US20030119013A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Bo
 ; APPLICANT: Hu, Wengli

APPLICANT: Tishkoff, Daniel
 APPLICANT: Zamudio, Carlos
 APPLICANT: Eroshkin, Alexey M
 APPLICANT: Lemieux, Sebastien M
 TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
 TITLE OF INVENTION: Methods of Use
 FILE REFERENCE: 10182-018-999
 CURRENT APPLICATION NUMBER: US/10/128,714
 CURRENT FILING DATE: 2002-04-23
 PRIOR APPLICATION NUMBER: US 60/285,697
 PRIOR FILING DATE: 2001-04-23
 PRIOR APPLICATION NUMBER: US 60/287,066
 PRIOR FILING DATE: 2001-04-27
 PRIOR APPLICATION NUMBER: US 60/295,890
 PRIOR FILING DATE: 2001-06-05
 PRIOR APPLICATION NUMBER: US 60/303,899
 PRIOR FILING DATE: 2001-07-09
 PRIOR APPLICATION NUMBER: US 60/316,362
 PRIOR FILING DATE: 2001-08-31
 NUMBER OF SEQ ID NOS: 8603
 SOFTWARE: Patent version 3.1
 SEQ ID NO 7139
 LENGTH: 2232
 TYPE: DNA
 ORGANISM: Aspergillus fumigatus
 US-10-128-714-7139

Alignment Scores:
 Pred. No.: 2,036-213 Length: 2232
 Score: 2172.50 Matches: 427
 Percent Similarity: 71.24% Conservativity: 108
 Best Local Similarity: 56.86% Mismatches: 171
 Query Match: 53.70% Indels: 45
 DB: 15 Gaps: 11

US-10-028-384-8 (1-774) x US-10-128-714-7139 (1-2232)

18 LeuileThrPheAlaileLeuLeuileAlaTrpLeuAlaGlyPheSerSerArgLeuPhe 37
 61 CTTCTCGGATTATTATCTGTTGATCCATTCGACGAGCAGCAGTGTCCGACTGCTTC 120
 38 AlaValileArgPheGluSerileleHisGluPheAspProTrpPheAsnTyArgAla 57
 121 AGCGTTATCGGTTCCGAGATATATCCAGGAAGTTGACCCGTTCACTCCGAGCA 180
 58 ThrAlaTyMetValGlnAsnGlyTrpTyAsnPheLeuAsnTrpPheAspGluArgAla 77
 181 ACAATACTTAGTACAGATGTTCTTATAGCTTTTGGATTGTTGATGACGGAACA 240
 78 TrpTyProLeuGlyArgileValGlyThrValTyProGlyLeuMetileThrSer 97
 241 TGGCATCTCTGGAGCTGTACCCGTTGACGATATATCCCGTCTCATGGTACGAGC 300
 98 GlyGlyileHisTrpLeuLeuHisValLeuAsnileProValHisileArgAspPheCys 117
 301 GCGGTGATCTACATATCTTGGATCTTCTATATCCCGTCTGATTTGCAACATCTGC 360
 118 ValPheLeuAlaProPheSerGlyLeuThrSerileSerThrTyLeuLeuThrLys 137
 361 GTCTACTCGGCGCAGATCTCCGCGCTGACTGCAATTTGGCAATGCTGTCATCTCC 420
 138 GluLeuTrp--SerAlaGlyAlaGlyLeuPheAlaAlaSerPheileAlaileValPro 156
 421 GAGATGTCCTTCCGATCTGAGGTCTTCTTGGAGAGCTTTTATGGGATCCCGCT 480
 157 GlyTyTrileSerArgSerValAlaGlySerTyAspAsnGluGlyleAlailePheAla 176
 481 GGTTACATCTCCGATCAGTTGTTGGAAGTACGATACGAAAGCAGATTTGCTTCTG 540
 177 LeuGlnPheThrTyPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
 541 CTTGTGTTACATCTTCTTATGATCAAGGCTGTCAAAATGGGTCTATCATGTGGGA 600

197 AlaAlaAlaLeuSerTyPheTyMetValSerAlaTrpGlyTyValPheile 216
 601 CCGCTGACCGGACTATTCTAGCGGTACATGGTGTGGGATGGGTATGTTCTTATT 660
 217 IleAsnLeuileProLeuHisValPheValLeuLeuileMetGlyArgTySerProArg 236
 661 ACGAACCTGATCCCGCTGCACTTTTGTCTTCTGTGCAATGGGTAGATACAGCACTCCG 720
 237 LeuLeuThrSerTyPheTyPheTyPheTyPheTyPheTyPheTyPheTyPheTyPhe 256
 721 ATCTACATTAGTATACCAATGATGATGCGGTGGGACTTTGGCTAGCATGCGATTCCC 780
 257 PheValGlyPheGlnProileArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
 781 TTCGTGCGATTTTTCCTTATCCGAAACAGCAGCAGCATGTCCGCTTTGGGTGTCTCCG 840
 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
 841 CTGCTTCAGCTGTGGCTTCGCGAGTTTGTCCGAGCTTCATCATCTCGGTCTCGGTGGT 900
 297 ArgLysLeu-----PheileValGlyGlyLeu---LeuValGlyValGly 310
 901 CAGAGACTTCTGACCGCCATGATCTTCATCATCTCGGTCTCGGTGGTGGT----- 954
 311 ValPheValAlaValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArg 330
 955 -----CTAGTTGTTCTGACTGTGAGCGGAGTATCGCTCTCTGGAGCGGCGCA 1002
 331 PheTySerLeuTrpAspThrGlyTyAlaLysileHisileProileleAlaSerVal 350
 1003 TTCTACTCTTGTGGGACATGCTATGCCAAATCCACATTCATCATCTCCATCATCTCC 1062
 351 SerGluHisGlnProTrpThrTrpPheSerPhePheAspLeuHisileLeuValCys 370
 1063 TCGGAACACACGCGCTTGGCCAGCTTCTTCTCGATCGAATCTCTGATCTGG 1122
 371 AlaPheProValGlyValTrpTyCysileLysGlnileAsnAspGluArgValPheVal 390
 1123 CTTTTCGCGGAGGTCTACATGCTTCGCTCGACCTCAAGGAGCATGCTCTCGTC 1182
 391 ValLeuTyAlaileSerAlaValTyPheAlaGlyValMetValArgLeuMetLeuThr 410
 1183 ATTATCTACTCGTCTTGGGACTACTTGGCGGTGTTATGTCGCACTATGCTGACC 1242
 411 LeuThrProValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPhe 430
 1243 TTGACCCCTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1302
 431 LeuGlnGluAspSerSerLysArgMetGlyThrAlaileSerAlaAlaThrGluValAsp 450
 1303 ATGGCGACT-----ACCTCCGACACACGCGCTCTGAAGCGAAA 1341
 451 GluAlaGluAspSerileGlyLysThrLeuTyAspLysAlaGlyLysLeuLysHis 470
 1342 ACGAATGAAGACTCGTCTTCAACACT----- 1368
 471 ArgThrLysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSerileVal 490
 1369 ---CTTGGCTCAGTTCGGAAGCCCAATGTTGAATCACTCCCATGTTTCTTAAGATTATA 1425
 491 IleLeuAla---ValLeuMetLeuLeuMetMetPheAlaValHisCysThrTrpValThr 509
 1426 GTGACGCGCTCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 1485
 510 SerAsnAlaTySerSerProSerileValLeuAlaPheHisAsnSerGlnAspGlySer 529
 1486 TCGAATGATACATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1542
 530 ArgAsnileLeuAspAspPheArgGluAlaTyTrpTrpLeuSerGlnAsnThrAlaAsp 549
 1543 CAATACATCATTCAGCATTCATCTGAGGCTTACTACTGCTTCTGCTGAGTAATCTCTCTAG 1602
 550 AspAlaArgValMetSerTrpTrpAspTyGlyTyGlnileAlaGlyMetAlaAsnArg 569

| | | | | |
|------|----|------|---|------|
| 1603 | Db | 1603 | AAAGCCAAATCATGTCATGGTGGATATGGGTATCAATCGGTGGCATGGCGGACCGC | 1662 |
| 570 | QY | 570 | ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla | 589 |
| 1663 | Db | 1663 | CCAACTCTGGTTGACAAACAACCTGGAACAACACCCATATTGCTACGGTTGGTAAGCGC | 1722 |
| 590 | QY | 590 | MetSerSerThrGluGluLysSerTyrgluIleMetThrSerLeuAspValAspTyrVal | 609 |
| 1723 | Db | 1723 | ATGAGCTCAGCGAGGAGTCAGCTACCCATCTCCGCAGCATGATGCGATTACGTCG | 1782 |
| 610 | QY | 610 | LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp | 629 |
| 1783 | Db | 1783 | CTGGTGGTGTGGTGGTCTGCTAGGTATTCTGGCGATCAGCATTAACAATTCATTATGC | 1842 |
| 630 | QY | 630 | MetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPheThr | 649 |
| 1843 | Db | 1843 | ATGGTCCGATCCCGAAGATATCGGCCGATGAGGTTAAAGAGCGGGACTTCCTTACT | 1902 |
| 650 | QY | 650 | AspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeuMet | 669 |
| 1903 | Db | 1903 | GCACGGCGTGAATATCGTGTGACGATGGAGCGACCCCAACTATGCGCACACGTTGTGG | 1962 |
| 670 | QY | 670 | TyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGly | 689 |
| 1963 | Db | 1963 | TATAAATGTCTATTACAAATTCAACTCTCTC-----TTCCGTCGGGC | 2007 |
| 690 | QY | 690 | -----TyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyr | 707 |
| 2008 | Db | 2008 | CAAGCTGTCGACGCGTCGTGGGTCAAAACITCCACAGAGAGCCCTCAGCTCTCTACA | 2067 |
| 708 | QY | 708 | LeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLysLysPro | 727 |
| 2069 | Db | 2069 | CTCGAAGAAGCTTTCACGAGCGAGACTGGATCATTCGTATCTCTACAAGGTCAAGGATCTT | 2127 |
| 728 | QY | 728 | HisGluPheAsnArgProSerLeuLysThrLysGluArgThrIleProAlaAsnPhe | 747 |
| 2128 | Db | 2128 | GACAACCTTGGCCGA-----GACCACACACCGCTGTGTCCCTTCGACAAAGGT | 2175 |
| 748 | QY | 748 | IleSerArgLysAsnSerLysArgArgLysGly | 758 |
| 2176 | Db | 2176 | CTCAAGAAACCGGAGTACAAGAGGAGGGA | 2208 |

RESULT 15

```

US-10-320-797-2305
; Sequence 2305, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2305
; LENGTH: 2466
; TYPE: DNA
; ORGANISM: Cryptococcus neoformans
US-10-320-797-2305

```

| Alignment Scores: | |
|---------------------|----------|
| Read No.: | 2 9e-207 |
| Score: | 2113.50 |
| Percent Similarity: | 70.48% |
| Local Similarity: | 54.48% |
| Query Match: | 52.24% |
| DB: | 16 |
| Length: | 2466 |
| Matches: | 407 |
| Conservative: | 182 |
| Mismatches: | 175 |
| Indels: | 43 |
| Gaps: | 10 |

US-10-028-384-8 (1-774) x US-10-320-797-2305 (1-2466)

| | | | |
|----|------|--|------|
| Qy | 4 | ThrProIyMetLeuAsnSerIySValAlaGlyTyrSerSerLeuileThrPheAlaIle | 23 |
| Db | 199 | TCTCCGTCGACTATCAACAACACG-----GAAAGCTGTGGCTTCATCATC | 246 |
| Qy | 24 | LeuLeuileAlaTrpLeuAlaGlyPheSerSerArgLeuPheAlaValIleArgPheGlu | 43 |
| Db | 247 | CTGGCGTTAATATGCGGTGCTGCGATTGGAAAGTCGATTGTTGCGGTGATCAGATTGAA | 306 |
| Qy | 44 | SerIleIleHisGluPheAspProTrpPheAsnTyrArgAlaThrAlaTyrMetValGln | 63 |
| Db | 307 | TCTGTCACTCCAGAAATTGACCCCTGGTTCATCTACCGAGCCTCGAAAGTTCITGTTAAC | 366 |
| Qy | 64 | AsnGlyTyrTrpAsnPheLeuAsnTrpPheAspGluArgAlaTrpTyrProLeuGlyArg | 83 |
| Db | 367 | RAAGGTTTCTACGAGTTCTCGAACTGCTTTGACCCCTCGCTTGGTACCTCTCGGCAGA | 426 |
| Qy | 84 | IleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyIleHisTrpLeu | 103 |
| Db | 427 | ACTGTGCGGTACCACGCTCTATCTGGCTTGATGTCACTCTGGACTGATTGGCATGCT | 486 |
| Qy | 104 | LeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeuAlaProIle | 123 |
| Db | 487 | CTTCGGGCAATCAATATGCCGTCGACATTCGCAATGTCGTGCTCTCTGCACCTGGA | 546 |
| Qy | 124 | PheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrySgluLeuTrpSerAlaGly | 143 |
| Db | 547 | TTTTCTGATTGATGCTGGCGGACTATCTTTTACCAGCTGAAATGCTACACCATCA | 606 |
| Qy | 144 | AlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSerArgSerVal | 163 |
| Db | 607 | GCTGTCTATTGGCGGCGCTTTCTATTGGCATTTGACCGGATACATCTCTCGATCTGTC | 666 |
| Qy | 164 | AlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThrTyrPheLeu | 183 |
| Db | 667 | CGCGGTTCTTATGACACAGAGCCATGCGCATCTCTCTTGATGAGCTGCTTCTACTCT | 726 |
| Qy | 184 | TrpValArgSerValIySThrGlySerValPheTrpSerAlaAlaAlaLeuSerTyr | 203 |
| Db | 727 | TGGATTAAAGCCGTCAAAACCGGTAGCTCATTTGGGTATGATCATCTGCTGCTTGTCTAC | 786 |
| Qy | 204 | PheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIleProLeuHis | 223 |
| Db | 787 | GGGTGGAATGGTTGTGTGATGGGTGTGTACGTTTTTCATCACCACATGATTCATTGCAC | 846 |
| Qy | 224 | ValPheValLeuLeuMetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThr | 243 |
| Db | 847 | GCCTTTGTTCTCATTGTCTATGGCAGGTTTCACACCGGCTTTATACCGCTTACTCTTCC | 906 |
| Qy | 244 | PheTyrIleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGlnProIle | 263 |
| Db | 907 | TGGTATGTCTATTGGAATATGCGCTCCATGTCAGGTCCCTTTGTGGAGTTCTCTCCCATC | 966 |
| Qy | 264 | ArgThrSerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThr | 283 |
| Db | 967 | CGAACCTCTGAGCACATGGCGGCTTGGGTGTTTTGCGTCTGTGACAGTCATCGGATTC | 1026 |
| Qy | 284 | LeuArgHisLeuGlnSerValLeuSerArgAsnGluPheArg-----LysLeuPhe | 300 |
| Db | 1027 | GTCGAAGTCGTCGACGACTGCTGTCGTCGCAAGCAATTCAGCTCTCTCAAAGCTTTT | 1086 |
| Qy | 301 | IleValGlyGlyLeuValGlyValGlyValPheValAlaValValLeuThrMet | 320 |
| Db | 1087 | GTCGTGGCGGTATTCGCTCAGT-----TTGCTGCCCTCGTCACCTTCACTTTC | 1137 |
| Qy | 321 | LeuGlyValValAlaProTrpSerGlyArgPheTyrSerLeuTrpAspThrGlyTrAla | 340 |
| Db | 1138 | TCTGGATGATCGCCCTTCGCTGGAAGATTTATCTCTTTGGGATCTGGCTATGCG | 1197 |
| Qy | 341 | LysIleHisIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpPheSer | 360 |
| Db | 1198 | AAGTCCATGATGATTTATTCCTTCGCTCCGATCCGAAACACAGGCCACCGGTGGCCCTCA | 1257 |

QY 361 PhePheAspLeuHisIleLeuValCysAlaPheProValGlyValTyrCysile 380
DB 1258 TTCTACTTTGACCTCGAAGGCTTATCTCTTTCCCTGCGGTGCTTCTGGGTTC 1317
QY 381 LysGlnIleAsnAspGluArgValPheValLeuTyrAlaIleSerAlaValTyrPhe 400
DB 1318 AAGGAGCTTCGGATGACGAGATCTTCATCATCATTTATGCGGTCTCAGTGCCTATTTT 1377
QY 401 AlaGlyValMetValArgLeuMetLeuThrLeuThrProValValCysMetLeuAlaGly 420
DB 1378 GCCGCTGTCATGTTGCGACTTATGCTGTATCATCACGCTGTGTCTGTGTTCTCCGCC 1437
QY 421 ValAlaPheSerGlyLeuLeuAspValPheLeuGlnGluAspSerSerLysArgMetGly 440
DB 1438 ATTGCGTTCCTCAAACTTCGAGGGGTATTTGAC----- 1473
QY 441 ThrAlaIleSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysLysThr 460
DB 1474 -----CCGTCATCCCGAAGCGAC-----GAGGAAGCTGGGAGCTCTCAGACG 1518
QY 461 LeuTyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAsp--- 479
DB 1519 CAGGTTGCTCCAAAGTCCAGGCGAAGAAGATGGCGCTGCCAACGCCAATAAGAGCGGG 1578
QY 480 -----ThrGlyValSerSerAsnLeuLysSerIle----- 489
DB 1579 TTCTCTTTACAGGTATTTGAGCGGCG---NAGTCGTCTCGGCATCTTTGGTCTCGAC 1635
QY 490 -----ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCys 505
DB 1636 ACTCGATTGCTGTGGTTTCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1695
QY 506 ThrTyrValThrSerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSer 525
DB 1696 ACATATGTGACTTCACAGCGGTATTTTCGCTTCAGTGTGATTCGATCGCGAAC--- 1752
QY 526 GlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTyrLeuSerGln 545
DB 1753 CCGGATGCTAGCCAAATATCATTTGATGATTTCCGAGAGGCTTACTACTGGATTGCGCAA 1812
QY 546 AsnThrAlaAspAlaArgValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGly 565
DB 1813 AACACCCCGACAGACAGCGTCATCATGCTCTGGTGGATTACGGCTACAGATCGCTGT 1872
QY 566 MetAlaAsnArgThrThrLeuValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeu 585
DB 1873 ATGGCTGATCGCCCGCCACCTTTGTTGATAAATACCTGGATTAACACCCACATTCGCCA 1932
QY 586 ValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAsp 605
DB 1933 GTTGTGAAGCCATGCTTCCACGAGATGTCGATATCTCTCTCTGAGGAGCATGAT 1992
QY 606 ValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsn 625
DB 1993 GTCCATTACGTTCTTGTGATCTTTGGGGCTTATTGGGCTACTCTGTTGAGCATATCAAC 2052
QY 626 LysPheLeuTyrMetValArgIleAlaGluGlyGluHisProLysAspIleLysGluSer 645
DB 2053 AAGTTTGTGGATGTTAGGATCTCACAGGATGATGCGCTGACAGAGTGCAGGAGTC 2112
QY 646 AspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeu 665
DB 2113 AACTCTCTCTCAAGAGGGAGTATGCTGTGATGACAGGGGCCACCCCTACTATGAAG 2172
QY 666 AsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArg 685
DB 2173 AACTCTCTCATGTACAAATGCTTACTACCGCTTCCCGAGCTT-----TATGGT 2223
QY 686 GlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeu 705
DB 2224 GGACACCCCGCTCAAGACAGGGTTGAGGCCAAATTTATCCCTCTCTCAAGTGTACTCTT 2283

QY 706 ThrTyrLeuGluGluAlaTyrThrThrGluHisTyrLeuValArgIleTyrArgValLys 725
DB 2284 GATACCTTTGACGAGCGTTCCATCTCCGAAATTTGGATCGTCAGGATCTACAAGGTCAAG 2343
QY 726 LysProHisGluPheAsnArg 732
DB 2344 AAGGAAGATCCCATTTGGACGA 2364

Search completed: December 15, 2004, 21:18:32
Job time : 1112.48 secs

Qy 590 MetSerSerThrGluGluLysSerTyrGluLeuMetThrSerLeuAspValAspTyrVal 609
 Db 2161 ATGCTTCAATGAACCGCTGCTTATGAATAATATGAAGCTTGATGATATGTA 2220
 Qy 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspIleAsnLysPheLeuTrp 629
 Db 2221 CTAATAATATTGGGGGTGAATTTGGATATCTTGGTGATGACATCAACAAATTCCTCTGG 2280
 Qy 630 MetValArgIleAlaGluGlyGluHisProLysPheAspIleLeuGluSerAspTyrPheThr 649
 Db 2281 ATGTTTGAATAGCAGAGAGAGAACATCCAAAGATATCAGGAAAGTGAATCTACCTCACC 2340
 Qy 650 AspArgGlyGluPheArgValAspAlaGluGlyValAlaProAlaLeuLeuAsnCysLeuMet 669
 Db 2341 A--CAAGGAGAGTTTCTGTAGACAAAGCTGGATC--CCAACCTTGTCTCAATTCCTCATG 2398
 Qy 670 TyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGlyProSerGly 689
 Db 2399 TATAGATGTCATATTAACCGGTGTGGGAATCGACCTTGATTTCCGAACACACCTGGA 2458
 Qy 690 TyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLeuGlu 709
 Db 2459 TTTGACCGGACACGCTAATGCTGAAATTTGGGAATAAGGATATTAATTCAGCATTTGGAA 2518
 Qy 710 GluAlaTyrThrThrGluHisTyrLeuValArgIleTyrArgValLysLysProHisGlu 729
 Db 2519 GAACATTTACATCGGAGCACTGGCTGTGTAGATATACAAAGTTAAAGAGCTAGAA--- 2575
 Qy 730 PheAsnArgProSerLeuLysTyrLysGluArgThr-----IleProProAlaAsn 746
 Db 2576 --AACGAGAGGCACTTGATCATATAACCTAGATTAACCAATATTGTACCAAAACAGAA 2632
 Qy 747 PheIleSerArgLysAsnSerLysArgGlyGlyTyrIleArgAsnArgProValVal 766
 Db 2633 TATTTATCAAAAAGACATCTAGAGAGAGCGTGCTGATAGGAAAGCTGACACG 2692
 Qy 767 ValLysGlyLysArgThrLeuLys 774
 Db 2693 AAGAAAGGCAAGACACCAACAG 2716

RESULT 3
 LOCUS AK012153 2669 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610524N02 product:hypothetical Oligosaccharyl transferase (OTase) STT3 subunit containing protein, full insert sequence.
 ACCESSION AK012153
 VERSION AK012153.1 GI:12848721
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2669)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Mishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tashiro, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.jp/) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATTCGAGTTAATTAATATCCCTCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
 Host: SOLR.
 FEATURES
 Location/Qualifiers
 1. 2669
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:2610524N02"
 /db_xref="taxon:10090"
 /clone="2610524N02"
 /tissue_type="whole body"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days embryo"
 misc_feature 36..2507
 /note="hypothetical Oligosaccharyl transferase (OTase) STT3 subunit containing protein (InterPro|IPR003674, evidence: InterPro)
 putative"
 polyA_signal 2654..2659
 /note="putative"
 polyA_site 2669


```

Db      2168  CTGCTTATGATAAAATGTCATACATGAGATTGGAGAAATCGACGATAGATTTCGCGAC 2227
Qy      686  yProSerGlyTrpAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuth 706
Db      2228  TCCCCAGGCTTGACCGAACACGTAATGCTGAGATTGGAATAAAGACATTAAATCAA 2287
Qy      706  rTyrlcugluclualatyrrthrghuHstTrrLeuValarglletyrrhrgvallyslsly 726
Db      2288  GCAATTTGGAGAAGCTTTATCATCAGAGCACTGCGCTTGTGAGATATATAGATGAAGC 2347
Qy      726  sProHisGluPheAsnArgProSerLeuLysThrLysGluArgThr-----llePr 743
Db      2348  ACCTGAC-----AACGGGAGACACTAGGTGCACAAACCTCGAGTCACCAACATCGTCC 2401
Qy      743  oProAlaAsnPhelSerArgLysAsnSerLysArgArgLysGlyTyrlleargAsnAr 763
Db      2402  CAACAGAGATATTGTCAAGAGAGACTACTAAAGGAAGCGTGCTGCTTAAATAATAA 2461
Qy      763  gProValValLysGlyLysArgThrLeuLys 774
Db      2462  GCTAGTGTTTAAGAAAGGCAAGAGACCTCTTAAG 2495

RESULT 4
LOCUS   BC028897
DEFINITION Mus musculus cDNA clone IMAGE:4168013, containing frame-shift errors.
VERSION BC028897
KEYWORDS HTC
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2656)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Kzywinski, M.I., Skalek, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2656)
DIRECTOR MGC Project.
DIRECT SUBMISSION
Submitted (04-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

```

```

Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 24 Row: f Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31560679
This clone has the following problem: frame shifted.

FEATURES
Location/Qualifiers
1..2656
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4168013"
/tissue_type="salivary gland, 10 week old female mouse"
/clone_lib="NCI CGAP_SG2"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.: 3,48e-238 Length: 2656
Score: 2200.00 Matches: 417
Percent Similarity: 73.82% Conservative: 113
Best Local Similarity: 58.08% Mismatches: 141
Query Match: 54.37% Indels: 47
DB: 3 Gaps: 7

US-10-028-384-8 (1-774) x BC028897 (1-2656)
Qy 17 SerLeuIleThrPheAlaIleLeuLeuLeuAlaGlyPheSerArgLeu 36
Db 65 ACATCTTAAGCTTCTCATCTCTCGATGCTGCTGTTATCTTTTCTACTCGCTT 124
Qy 37 PheAlaValIleArgPheGluSerIleIleHicGluPheAspProTrpPheAsnTrpArg 56
Db 125 TTCTGCTGCTGAGATTGAAGATGTCATCCATGAGTTGATCCGACTTTTAATATCGG 184
Qy 57 AlaThrAlaTyrrMetValGlnAsnGlyTrpTyrrAsnPheLeuAsnTrpPheAspGluArg 76
Db 185 ACTACCGGTTTCGGCTGAGGAGGGGTTTATAAATTCATTAATCGATTGATGACCG 244
Qy 77 AlaTrpTrpProLeuGlyArgIleValGlyGlyThrValTyrrProGlyLeuMetIleThr 96
Db 245 GCTTGGTACCTTTGGGCCGGAATCATGGAGGAAACAATTTACCACCGGTTTAATGATCCT 304
Qy 97 SerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIle 116
Db 305 TCTGCTGCATCTACCATGATCTCCATCTTCTTCCATATCATCTATTGACATTCGGAATGC 364
Qy 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrrLeuLeuThr 136
Db 365 TGTGTTTCTCGGCCCACTTTCTCTCTTTCACCCACCATCGTACCTACCTTACC 424
Qy 137 LysGluLeuTrpSerAlaGlyValGlyLeuPheAlaAlaSerPheIleAlaIleValPro 156
Db 425 AAAGAGCTCAAGGATGAGAGGCTGGGCTTCTTGTCTGCTGCCATGATGCTGTAGTTCCT 484
Qy 157 GlyTyrlleSerArgSerValAlaGlySerTyrrAspAsnGluGlyIleAlaIlePheAla 176
Db 485 GGGTATATTTCTCGATCTGTAGTGGCTCTCTATGATAATGAAGAAATTCGTATCTTTTC 544
Qy 177 LeuGlnPheThrTyrrPheLeuTrpValArgSerValLysThrGlySerValPheTrpSer 196
Db 545 ATGCTGCTTACTTACTACATGATGATGATCAAGGAGTGAAGACTGTTCCATCTATTGGGCT 604
Qy 197 AlaAlaAlaLeuSerTyrrPheTyrrMetValSerAlaTrpGlyGlyTyrrValPheIle 216

```


QY 471 gThryshisaspalagInaspthr-----GlyValSerSerAsnLeuIysSerI 489
 Db 1455 ~~~~~-AAGCAACAGGATTCTTACCTTATTAAGAAATGAGGTGGCAGTGG 1501
 QY 489 eValileLeuAlaValLeuMetMetMetPheAlaValHisCysThrTIPValrh 509
 Db 1502 GATGATACCTGGTCTGCTTCTTCATCACCCTACACGCTTCATTCGACITGGTGAC 1561
 QY 509 rSerAsuAlaTyrrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspIysE 529
 Db 1562 CAGTGAAGCCATTCTTCCTCCATGCTTCGTCTGCTGCTGGTGGG---GATGGCAG 1618
 QY 529 rArgAsnIleLeuAspAspPheArgGluAlaTyrrTIPLeuSerGlnAsnThrAlaAs 549
 Db 1619 TAGGATCATTTTTCATGACTTCCGAGAAGCGTATTATTGGCTCCGTCACAACTACCCAGA 1678
 QY 549 pAspAlaArgValMetSerTIPTrpAspTyrrGlyTyrrGlnIleAlaGlyMetAlaAsnAr 569
 Db 1679 GGATGCAAAAGTCATGTCATGGTGGGATTATGCTACCAAAATTACTGCAATGGCAAAATCG 1738
 QY 569 gThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyIysAl 589
 Db 1739 GACAATTTTAGTGACAATAACACATGGAATAATACCATAATTCTCCGAGTAGGCAGCG 1798
 QY 589 aVetSerSerThrGluGluIysSerTyrrGluIleMetThrSerLeuAspValAspTyrrVa 609
 Db 1799 AATGGCATCCACAGAAAGAGCCTATGAATCATGAGGAGCTTGATGTGACGTATGT 1858
 QY 609 lLeuValIlePheGlyGlyValIleGlyTyrrSerGlyAspAspIleAsnIysPheLeuTr 629
 Db 1859 GCTTGTCATTTTGGAGCGCTTACTGGTATTCTTCGGATGATATCAACAGTTCTTGG 1918
 QY 629 pMetValArgIle---AlaGluGlyGluHisProLysAspIleIysGlySerAspTyrrPh 648
 Db 1919 GATGTCGGGATTGGAGAGACACAGAGACAGAGACACATTAAGGAGAATGACTACTTA 1978
 QY 648 eThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuAsnCysLe 668
 Db 1979 TACTCTACTCGGGAATTCGGTGTGATCGTGGGTTCTCCGGTCTCTCACTGCT 2038
 QY 668 uMetTyrrIysLeuSerTyrrArgPheGlyGluLeuIysLeuAspTyrrArgIlyProSe 688
 Db 2039 TATGTACAAATGTTTACTACCGCTTGGCAGGCTACACAGAACCCAGCGTCCACC 2098
 QY 688 rGlyTyrrAspArgThrArgAsnAlaValIleGlyAsnIysAspPheAspLeuThrTyrrle 708
 Db 2099 AGGCTTTCACCGTGTTCGAATGCTGAGATTGGTATAAAGACTTTTGACCTTGATGCT 2158
 QY 708 uGluGluAlaTyrrThrThrGluHisTrpLeuValArgIleTyrrArgValIys 725
 Db 2159 GGAGAGAGCGTATACACAGAACTGGCTAGTCAGGATATACAGGTAAAG 2210
 RESULT 7
 AK077877 3643 bp mRNA linear HTC 03-APR-2004
 LOCUS Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
 DEFINITION enriched library, clone.6030401L06 product:OLIGOSACCHARYL
 TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
 homolog [Mus musculus], full insert sequence.
 ACCESSION AK077877
 VERSION AK077877.1 GI:26097543
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 REFERENCE
 AUTHORS Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,H., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Chara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multipillar sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 3643)
 AUTHORS Adachi,S., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
 Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayatsu,N., Hiramoto,K., Hirao,T., Hirozane,T.,
 Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
 Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
 Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
 Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,
 Yuramatsu,M. and Hayashizaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
 URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9216,
 Fax:81-45-503-9216)
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.jp/
 URL:http://fantom.gsc.riken.jp/
 FEATURES
 source
 1. .3643
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CS7BL/6J"
 /cd_xref="FANTOM_DB:6030401L06"
 /db_xref="taxon:10090"
 /clone="6030401L06"
 /sex="male"
 /tissue type="testis"
 /clone lib="RIKEN full-length enriched mouse cDNA library"
 /dev stage="13 days embryo"
 misc_feature 93..2209

/note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus] (SWISSPROT)P46978, evidence: FASTY, 99.7%ID, 100%length, match=2115) putative"

ORIGIN

Alignment Scores:

Pred. No.: 1,326-234 Length: 3643
 Score: 2170.50 Matches: 415
 Percent Similarity: 73.64% Conservative: 113
 Best Local Similarity: 57.86% Mismatches: 144
 Query Match: 53.65% Indels: 46
 DB: 3 Gaps: 8

US-10-028-384-8 (1-774) x AK077877 (1-3643)

```

QY 17 SerLeuIleThrPheAlaIleLeuLeuIleAlaIleLeuAlaGlyPheSerSerArgLeu 36
DB 138 ACACCTCTAAAGCTTCTCATCTCTGTCGATGGCTGCTGTTATCTTTTCTACTCGTCT 197
QY 37 PheAlaValIleArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArg 56
DB 198 TTTGCTGTCTGAGATTTTGAAGTGTCTATCCATGAGTTTGATCCGTACTTTAATATTCGG 257
QY 57 AlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyrPheAspGluArg 76
DB 258 ACTACCGGTTTCTGGCTGAGGAGGGGTTTATAAATTCATCTGTTGATGACCGG 317
QY 77 AlaTyrTyrProLeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThr 96
DB 318 GCTTGCTACCTTTTGGCGGATCATCTGGAGAACATTTACCAGGTTTAAATGATCACT 377
QY 97 SerGlyGlyIleHisThrPheLeuHisValLeuAsnIleProValHisIleArgAspIle 116
DB 378 TCTGCTGCAATCTACCATGCTACTCTCTTCTCCATATCACCATTGACATTCGGAGTGC 437
QY 117 CysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThr 136
DB 438 TGTGTTTCTGCGCCCGCTTTTCTCTCTTTCACCCACCATCGTTAGCTACCATCTTACC 497
QY 137 LysGluLeuTyrSerAlaGlyAlaGlyLeuPheAlaLeuSerPheIleAlaIleValPro 156
DB 498 AAGAGCTCAAGATGCAGAGCTGGGCTTCTGCTGTCGATGATGCTGTGATTCCT 557
QY 157 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAla 176
DB 558 GGGTATATTTCTCGATCTGTAGCTGGCTCCTATGATAATGAAGAAATTCATCTTTTGC 617
QY 177 LeuGlnPheThrTyrPheLeuTyrValArgSerValIleThrGlySerValPheThrSer 196
DB 618 ATGCTGCTTACTTACTACATGTGGATCAAGGAGTGAAGACTGTTCCATCTATGGGCT 677
QY 197 AlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrGlyGlyTyrValPheIle 216
DB 678 GCCAAGTGTGCGCTCGCTTATTTCTACATGGTCTCTTCATGGGAGGCTATGTGTTCTCT 737
QY 217 IleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyValGlyTyrSerProArg 236
DB 738 ATCAACTTGTATTCCTTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 797
QY 237 LeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGlnIlePro 256
DB 798 ATCTAGCTAGCTTACTGTACTGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857
QY 257 PheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeuGlyValPheVal 276
DB 858 TTTGTTGGTTTCCAGCCCGCTCTTTCATCAGAACACATGGGAGGCTTTGGAGTGTTCGT 917
QY 277 LeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSerArgAsnGluPhe 296
DB 918 CTCTGTGATCATCATCTCTTCTAGATTACCTTGGCGAGCAAGTTGAATCCACAGCAATTC 977

```

```

QY 297 ArgLysLeuPhe-----IleValGlyGlyLeuLeuValGlyValGlyVal 311
DB 978 GAAGTTCCTTTCCGAGGTGTTATCTCCCTGGTTGGCTTTGCTCTCTCACTGTGGGA--- 1034
QY 312 PheValAlaValValLeuThrMetLeuGlyValValAlaProThrSerGlyArgPhe 331
DB 1035 -----GCTCTCTCTCATGCTAACA-----GGAAAAATTTCTCCCTGGACAGGCGTTTC 1082
QY 332 TyrSerLeuTyrAspThrGlyTyrAlaIleHisIleProIleAlaSerValSer 351
DB 1083 TACTCTCTCTGGATCCCTCTTATGCTAAGAATAACATCCCATTTATGCTATCTGTTCT 1142
QY 352 GluHisGlnProThrThrTyrPheSerPhePhePheAsnLeuHisIleLeuValCysAla 371
DB 1143 GAGCACAGCCCAACACCTGCTCTCTATTTATTTATCTACAGCTCTTGTCTTCATG 1202
QY 372 PheProValGlyValTyrTyrCysIleGlnIleAsnAspGluArgValPheValVal 391
DB 1203 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTCATGCTCGGATTTTATCATC 1262
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
DB 1263 ATGTATGCTGTGACGAGCATGCTCTTTCAGCTGTAATGTGGTCTTAATGCTGTGATTG 1322
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431
DB 1323 GCACCTGTTATGTGCTCTTCTGGCATTTGCTGCTTTCCAGGCTGCTGCCATATATG 1382
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
DB 1383 AAAAATCTGGCAATAAGTCGCCCA----- 1406
QY 452 AlaGluAspSerIleGluLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
DB 1407 -----CACAAAGAGCAAG----- 1421
QY 472 ThrLysHisAspAlaGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
DB 1422 -----AGCAACAGGATCTACTTACCTATTAAAGATGAGGTGGGAGTGGG 1469
QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
DB 1470 ATGATACTGGTCATGGCTTTCTTCTCATCACCTACACGTTTCACTCGACTGGGTGACC 1529
QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 529
DB 1530 AGTGAAGCTATTTCTTCTCCCTCCATTTCTGCTGCTCGTGGTGGG---GATGGCAGT 1586
QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTyrTrpLeuSerGlnAsnThrAlaAsp 549
DB 1587 AGGATCATTTTGTATGACTTCCGAGAGCGTATTATTGGCTCCGTCACATACTCCAGAG 1646
QY 550 AspAlaArgValMetSerTrpTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
DB 1647 GATGCAAAAGTCATGTCATGTGGGATTTGCTTACCAATTTACTGCAATGGCAATCGG 1706
QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
DB 1707 AA-ATTTAGTGGACAATAACACATGAATAATATACCCTATTTCTCGAGTAGGCGGCA 1765
QY 590 MetSerSerThrGluGluLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
DB 1766 ATGGCATCCACAGAGAGAAAGCCCTATGAATCATGAGGAGCTTGATGTCAGCTATGTG 1825
QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
DB 1826 CTTGTCATTTTGGAGGCTTACTGGGTATTTCTTCGGATGATATCAACAAGTTTCTTTGG 1885
QY 630 MetValArgIle---AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 1886 ATGGTCCCGGATTTGGAGGAGACAGACAGAGAACACATTAAAGGAGATGACTACTAT 1945
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668

```

```

Db 1946 ACTCCTACTGGGAATTCCTGTGTGATCGAGGGTCTCCGGTGTGCTCAACGCTT 2005
Qy 669 MetTyrlsLeuSerTyTyArgPheGlyGluLeuLysLeuAspTyArgGlyProSer 688
Db 2006 ATGTACAAATGTCTTACTACCGCTTTGGCAGGCTCTACACAGAGCCAGCGTCCACCA 2065
Qy 689 GlyTyArgPheArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyLeu 708
Db 2066 GCCTTTGACCGTGTTCGAATGCTGAGATTGGTAATAAGACCTTTGAGCTTGCATGCTG 2125
Qy 709 GluGluAlaTyThrThrGluHisTrpLeuValArgIleTyArgValLys 725
Db 2126 GAGGAAGCGTATACACAGAACACTGGCTAGTCAGGATATACAGGTAAG 2176

RESULT 8
AY418284 2118 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY418284
VERSION AY418284.1 GI:39774244
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Clark A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS Clark A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..2118
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2118
/gene="ITM1"
/locus_tag="HCM6506"

gene
Alignment Scores:
Pred. No.: 1,01e-223 Length: 2118
Score: 2071.50 Matches: 393
Percent Similarity: 72.12% Conservative: 101
Best Local Similarity: 57.37% Mismatches: 140
Query Match: 51.20% Indels: 51
DB: 9 Gaps: 8

US-10-028-384-8 (1-774) x AY418284 (1-2118)
Qy 52 TrpPheAsnTyArgAlaThrAlaTyMetValGlnAsnGlyTyTrpTyArgAsnPheLeuAsn 71
Db 151 TACTTTAATATCGGACTACAGGTTCTCGCTGGAGGGGGTTTATAAATTCCTAATAC 210
Qy 72 TrpPheAspGluArgAlaTyTrpProLeuGlyValArgIleValGlyThrValTyPro 91
Db 211 TGGTTTGTATGACCGGCTGTGATCCCTTTGGGAGCAATCATTTGGAGGACAAATTTACCCA 270

```

```

Qy 92 GlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProVal 111
Db 271 GGTTTAATGATCACTCTCTGCAATCTACCATGTACTCCATTTTTCACATCACCATT 330
Qy 112 HisIleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSer 131
Db 331 GACATTGGGAATGTCGTGTCTCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 390
Qy 132 ThrTyLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPhe 151
Db 391 ACGTACCACTTACCAAGAGCTCAAGGATGACGGGCTGGGCTTCTTGTCTGCTGCCATG 450
Qy 152 IleAlaIleValProGlyTyTrpIleSerArgSerValAlaGlySerTyArgAsnGluGly 171
Db 451 ATTGCTGTAGTTCCTGGATATATCTCCGATCTGGGCTGGCTCTCTATGATATGAAGAGG 510
Qy 172 IleAlaIlePheAlaLeuGlnPheThrTyPheLeuTrpValArgSerValIleThrGly 191
Db 511 ATTGCACTTTTTCATGCTACTACCTACTACATGTGGATCAAGGACAGTAAAGACTGGT 570
Qy 192 SerValPheTrpSerAlaAlaAlaLeuSerTyTrpMetValSerAlaTrpGly 211
Db 571 TCCATCTGTGGGACGCTAAGTGTCCCTTGTCTTATTTCTACATGGTCTCTGTCATGGGA 630
Qy 212 GlyTyValPheIleLeuAsnLeuIleProLeuHisValPheValLeuLeuIleMetGly 231
Db 631 GGTATGCTCTCTGATCACTTAATTCCTCTCCACGCTCTCTGCTGTGATGCTCACAGGC 690
Qy 232 ArgTySerProArgLeuLeuThrSerTyTrpSerThrPheTyTrpIleLeuGlyLeuPhe 251
Db 691 CGTTTCTCTCACCGGATCTATGTGGCTACTGCTGTCTCTCTCTCTCTCTCTCTCTCTCT 750
Qy 252 SerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 271
Db 751 TCTATGCAGATCTCTTTGTGGTTTCCAGCGCTGCTCTCTCTCATCAGAGCACATGGCAGCC 810
Qy 272 LeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeu 291
Db 811 TTTGGGGTCTTTGGTCTCTGCCAGATCCATGCTTTGTGGATTACCTGGCGGACGAGTTG 870
Qy 292 SerArgAsnGluPheArgLysLeuPhe-----IleValGlyGlyLeuLeu 306
Db 871 AATCCACAAATTTGAAGTCTTTTCCGAGCGCTCATCTCTCTGTAGGCTTTGTCTCT 930
Qy 307 ValGlyValGlyValPheValAlaValValLeuThrMetLeuGlyValValAlaPro 326
Db 931 CTCACCGTGGGA-----GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 975
Qy 327 TrpSerGlyArgPheTySerLeuTrpAspThrGlyTyAlaLysIleHisIleProIle 346
Db 976 TGGACGGGCGTTTCTACTCGCTGCTGATCCCTCTTATGCTAAGAACAAACATCCCCATC 1035
Qy 347 IleAlaSerValSerGluHisGlnProThrTrpPheSerPhePhePheAspLeuHis 366
Db 1036 ATTGCTTCTGTCTGTGAGCATCAGCCCAACCTGGTCTCTCTCTCTCTCTCTCTCTCTCT 1095
Qy 367 IleLeuValCysAlaPheProValGlyValTyTrpTyCysIleLysGlnIleAsnAspGlu 386
Db 1096 CTCCTCGTCTCATGTTTCCAGTTCGCTCTATTACTGCTTTAGCAACCTGCTCTGTAGTGC 1155
Qy 387 ArgValPheValValLeuTyAlaIleSerAlaValTyPheAlaGlyValMetValArg 406
Db 1156 CGGATTTTATCATGATGTTGGTGTGACGAGCATGTTCTTTTCTGCTGTAAATGGTGGT 1215
Qy 407 LeuMetIleThrLeuTrpProValValCysMetLeuAlaGlyValAlaPheSerGlyLeu 426
Db 1216 CTAATGCTAGTGTGGCACCTGTTATGTGCACTTCTCTCTGCAATGGAGTCTCCAGGTG 1275
Qy 427 LeuAspValPheLeuGlnLysAspSerSerLysArgMetGlyThrAlaIleSerAlaAla 446
Db 1276 CTGTCCACATACATGAAGAATCTGGACATAAGTCTGTCCA-----1314

```


1763 CCATATTTCTCGAGTAGGCGAGCAATGGCATCCACAGAAAGAAAGGCTATGAATC 1822

601 MetThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSer 620

1823 ATGAGGAGCTTGTAGCTATGCTTGTATTTTGTGAGGCTTACTGGTATCT 1882

621 GlyAspAspIleAsnLeuPheLeuTyrMetValArgIle---AlaGluGlyGluHisPro 639

1883 TCGGATGATATCAACAAGTTCTTTGGATGCTCGGATTTGGAGGAGCAGACAGCA 1942

640 LysAspIleLeuGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGlu 659

1943 AGACACATTAGAGAGATGACTACTTCTCTACTGGGAAATTCCTGTTGATCGTGAG 2002

660 GlyAlaProAlaLeuLeuAsnCysLeuMetTyrIleLeuSerTyrTyrArgPheGlyGlu 679

2003 GGTCTCGGCTGCTCTCAACTGCTTATGTACAAAATGTTACTACCGCTTTGGCGAG 2062

680 LeuLysLeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGly 699

2063 GTCTACACAGAGCAAGGCTCCACAGGCTTTCACCGTGTTCGAAATGCTGAGATGGT 2122

700 AsnLysAspPheAspLeuThrTyrLeuGluGluAlaTyrThrGluHisTyrLeuVal 719

2123 AATAAAGACTTTGAGCTTGATGCTCTGGAGGAGCGTATACACAGAACACTGGCTAGTC 2182

720 ArgIleTyrArgValLys 725

2183 AGGATATACAGGTTAAG 2200

RESULT 10

AY310157

LOCUS Rattus norvegicus Acl1573 mRNA, complete cds. HTC 07-JUL-2003

DEFINITION Rattus norvegicus Acl1573 mRNA, complete cds.

ACCESSION AY310157

VERSION AY310157.1 GI:32264636

KEYWORDS HTC.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 2334)

AUTHORS Xu, C.S., Li, W.Q., Li, Y.C., Han, H.P., Wang, G.P., Chai, L.Q., Yuan, J.Y., Yang, K.J., Yan, H.M., Chang, C.F., Zhao, L.F., Ma, H., Wang, L., Wang, S.F., Shi, J.B., Rahman, S., Wang, Q.N. and Zhang, J.B.

TITLE Liver regeneration after PH

REFERENCE 2 (bases 1 to 2334)

AUTHORS Xu, C.S., Li, W.Q., Li, Y.C., Han, H.P., Wang, G.P., Chai, L.Q., Yuan, J.Y., Yang, K.J., Yan, H.M., Chang, C.F., Zhao, L.F., Ma, H., Wang, L., Wang, S.F., Shi, J.B., Rahman, S., Wang, Q.N. and Zhang, J.B.

TITLE Direct Submision

JOURNAL Submitted (29-MAY-2003) Henan Bioengineering Key Lab, Henan Normal University, No. 148 Jianshe Road, Xinxiang City, Henan 453002, P.R. China

FEATURES

source

1..2334

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

232..2334

/note="liver regeneration related protein LRRG109"

/codon_start=1

/product="Acl1573"

/protein_id="AAP78765.1"

/db_xref="GI:32264637"

/translation="MITAGLHWIINTVIRHRCVCFVLAFTFSLTSTISFLTR ELWNGAGLLAACFTAIIPGYSIRSVAGSFDNEGIAIFALQFTYIWMHSGVVLCPFA HSPGPFSSGHSHTISLVAALPSGALVHVARPELYRVSAGWYVPIINLIPLHVFV LLLMORYSKRYVIGVFALLQAYAFLOVLRDLTKQEPQTLFFLCVSLAAGAVLSVIY LTYTGMQKVTYFOLLHQLCLNISLRHGLSSLIYIFLCVPSQOAYGSKISTMKEYLAG

CDS

ORIGIN

Alignment Scores:

Prod. No.: 1,32e-221 Length: 2334

Score: 2053.50 Matches: 416

Percent Similarity: 65.66% Conservative: 64

Best Local Similarity: 56.91% Mismatches: 93

Query Match: 50.75% Indels: 159

DB: 3 Gaps: 9

US-10-028-384-8 (1-774) x AY310157 (1-2334)

QY 53 PheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTyr 72

Db 109 TTTAACTATAGATCAACTCATCTTGCATCTCATGGTTCATGAGTCTCTTAAATGG 168

QY 73 PheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGlyThrValTyrProGly 92

Db 169 TTTGATGAAGAGAGCATGTCACCCACTAGGAAGATCTGGTGGCACCCTTACCAGGA 228

QY 93 LeuMetIleThrSerGlyGlyIleHisTyrLeuLeuHisValLeuAsnIleProValHis 112

Db 229 TTGATGATACAGCTGGCTTATTCTTGGATTAAATACATATGTAACATACAGTTCAC 288

QY 113 IleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThr 132

Db 289 ATCAGAGATGATGTGTCTCTTGCACCAACTTTTAGCGGCTTACATCTATATCTACG 348

QY 133 TyrLeuLeuThrLysGluLeuTyrSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIle 152

Db 349 TTCTGCTTAACATAGAACATTGGAACAGGAGCAGGCTCTTAGCTGCTCTCAT 408

QY 153 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIle 172

Db 409 GCTATTGTACCGGGTACATATCTCGTTCGTAGCGGATCTCTTGATACGAAGCAT 468

QY 173 AlaIlePheAlaLeuGlnPheThrTyrPheLeuTyrValArgSer----- 187

Db 469 CGGATTTTCGACATCTCATCTACTACTATTAGGGCCATCTCGGGTGTGCTCTCG 528

QY 188 -----ValLysThrGlySerValPheThrSerAla----- 197

Db 529 TTTCTGCGCCATCGCTCTCCAGGCGCTTTGGTTCTCTCAGGCGCATTCATCACCATCC 588

QY 198 -----AlaAlaAlaLeu-----SerTyrPheTyr 205

Db 589 TTGTGGCGCGTGTCTCTCTAGTGGGCGCTGTGATGTCATGTTCTCGGCTGAGCTCTAC 648

QY 206 MetValSerAlaTyrGlyTyrValPheIleIleAsnLeuIleProLeuHisValPhe 225

Db 649 AGGCTCTCTGCTGGGAGGTTATGCTTTCATCATCACTCATCTCTCCATGTTT 708

QY 226 ValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyr 245

Db 709 GTGCTGCTCTCTGATGCAGAGGTACAGCAGAGAGTCTAC----- 747

QY 246 IleLeuGlyLeuLeuPheSerMetGlnIleProPheValGlyPheGlnProIleArgThr 265

Db 747 ----- 747

QY 266 SerGluHisMetAlaAlaLeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArg 285

Db 748 -----ATAGGTGTCTTTTCGCTGCTGCAGGCTTATGCGTTTCTCGAC 789

QY 286 HisLeuGlnSerValLeuSerArgAsnGluPheArgLysLeuPheIleValGlyLeu 305

QY 52 TrpPheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsn 71
DB 151 TACTTTAATTATCGGACTACCGGTTCTGGCTGAGAGGGGTTTATATATTCCTCAATAC 210
QY 72 TrpPheAsnGluAtrAlaTyrTrpProLeuGlyArgIleValGlyThrValTyrPro 91
DB 211 TGGTTTGAACCGGGCTGGTACCCCTTTGGCCGCAATCATTTGAGGAACAATTTACCCA 270
QY 92 GlyLeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProVal 111
DB 271 GGTTTAATGATCACCTTCGTGCAATCTACCATGTACTCCATTTCTCCATATCCATATT 330
QY 112 HistLeuArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSer 131
DB 331 GACATTCGGAATGCTGTGTTTCTGCGCCCATCTTTCTCTCTTTCACACCACCATCGTT 390
QY 132 ThrTyrLeuLeuThrLysGluLeuTrpSerAlaGlyValGlyLeuPheAlaAlaSerPhe 151
DB 391 ACGTACCACCTTACCAAGAGCTCAAGATGCAGAGCTGGGCTTCTGCTGCTGCTGCTG 450
QY 152 IleAlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGly 171
DB 451 ATTCTGTAGTTCCTGGGTATTTCTCGATCTCTAGCTGCTGCTGCTGCTGCTGCTGCT 510
QY 172 IleAlaIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGly 191
DB 511 ATTCTGTATCTTTCGATGCTGCTTACCTTACATGTGGATCAAGGAGGAGGAGCTGGT 570
QY 192 SerValPheTrpSerAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrGly 211
DB 571 TCCATCATTTGGGCTGCAAGTGTGCTGCTTATTTCTACATGCTCTCTTCATGGGA 630
QY 212 GlyTyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGly 231
DB 631 GCCTATGTTGCTTGTGATCACTTATCTCTACATGCTCTGCTGCTGCTGCTGCTGCT 690
QY 232 ArgTyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPhe 251
DB 691 CGTTTCTCACCGATCTAGTACCTACTGATCTGTTACTGCTGGCCACCATCTT 750
QY 252 SerMetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAla 271
DB 751 TCTATGAGATTTCTTGTGTGTTTCCAGCCGCTCTTTCATCAGAACACATGGCAGCC 810
QY 272 LeuGlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeu 291
DB 811 TTTGGAGTGTGCTCTCTGTCAGATCCATGCTTTGATGATTTACTGCTGCGCAAGTTG 870
QY 292 SerArgAsnGluPheArgLysLeuPheIleValGlyLeuLeuValGlyValGlyVal 311
DB 871 AATCCACAGCAATTCGAAGTCTTTTCCGGAGTGTATCTCCCTGCTGGC----- 921
QY 312 PheValAlaValValValLeuThrMetLeuGlyValValAlaProTrpSerGlyArgPhe 331
DB 922 TTTGCTCTCTACTGTGGAGCTCTCTCTCGA-AAAAATTTCTCCCTGAGCAGGGCGTTTC 980
QY 332 TyrSerLeuTrpAspThrGlyTyrAlaLysIleHisIleProIleIleAlaSerValSer 351
DB 981 TACTCTCTGCTGATCCCTCTTATGCTAAGATAACATTCCTCATATTGCTGCTTCT 1040
QY 352 GluHisGlnProThrThrTrpPheSerPhePheAspLeuHisIleLeuValCysAla 371
DB 1041 GAGCACAGCCCAACCTGGTCTCTCTACTATTTTGGATCTACAGCTCTCTCTCTCATG 1100
QY 372 PheProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValVal 391
DB 1101 TTTCCAGTTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCTCGGATTTTATCATC 1160
QY 392 LeuTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeu 411
DB 1161 ATGTATGGTGTGACACGATGACTTTTTCAGCTGTAATGCTGCTGCTGCTGCTGCTG 1220
QY 412 ThrProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeu 431

DB 1221 GCACCTGTTATGGCAATCTTTCTGGCAATGGTGTTCCTCCAGTGTCTGCCACATATATG 1280
QY 432 GlnGluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGlu 451
DB 1281 AAAAATCTGCACATAAGTCGCCCA----- 1304
QY 452 AlaGluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArg 471
DB 1305 -----GACAAGAGCAAG----- 1319
QY 472 ThrLysHisAspAlaGlnGlnAspThr-----GlyValSerSerAsnLeuLysSerIle 489
DB 1320 -----AAGCAACAGGATCTACTTACCTTAAAGAAATGAGTGGCGAGTGGG 1367
QY 490 ValIleLeuAlaValLeuMetLeuMetMetPheAlaValHisCysThrTrpValThr 509
DB 1368 ATGATACTGGTCATGGCTTTTCTCTCCCTCCATTTGCTGCTGCTGCTGCTGCTGCTG 529
QY 510 SerAsnAlaTyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySer 549
DB 1428 AGTGAAGCCTATCTTCTCCCTCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1484
QY 530 ArgAsnIleLeuAspAspPheArgGluAlaTyrTrpLeuSerGlnAsnThrAlaAsp 549
DB 1485 AGGATCATTTTTGATGACTTCCGAGAAGCGTATTATTGGCTCCGTCACATATCCAGAG 1544
QY 550 AspAlaAatGValMetSerTrpTrpAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArg 569
DB 1545 GATGCAAAAGTCATGCTGTTGGGATTTATGCTACCAATTTACTGCAATGGCAATCGG 1604
QY 570 ThrThrLeuValAspAsnAsnThrTrpAsnAsnSerHisIleAlaLeuValGlyLysAla 589
DB 1605 ACAATTTAGTGCACATAACACATGGAATAATATCCCATATTTCTCGAGTAGGCGAGCA 1664
QY 590 MetSerSerThrGluGlyLysSerTyrGluIleMetThrSerLeuAspValAspTyrVal 609
DB 1665 ATGGCATCCACAGAAAGAAAGCCATGAAATCATGAGGAGCTGTGATGTGCTGCTGCTG 1724
QY 610 LeuValIlePheGlyGlyValIleGlyTyrSerGlyAspAspIleAsnLysPheLeuTrp 629
DB 1725 CTGTGCAATTTTGGAGGCTTACTGGGTATTTCTTCGGATGATATCAACAAGTTCTTTGG 1784
QY 630 MetValArgIle---AlaGluGlyGluHisProLysAspIleLysGluSerAspTyrPhe 648
DB 1785 ATGTCTCCGATTTGGAGGAGCAGACAGACAGACACATTAAGGAGATGACTACTAT 1844
QY 649 ThrAspArgGlyGluPheArgValAspAlaGluGlyAlaProAlaLeuLeuAsnCysLeu 668
DB 1845 ACTCTACTGCGGAAATTCGGTGTGATGCGTGGGTTCTCCGGTGTGCTCACTGCTT 1904
QY 669 MetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLysLeuAspTyrArgGly-Prose 688
DB 1905 ATGTACAAAATGTCTTACTACCGCTTTGGGAGCTCTACACAGAA-----GCCTCCACC 1958
QY 688 rGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLysAspPheAspLeuThrTyrLe 708
DB 1959 AGGCTTTGACCGTGTTCGAAATGCTGAGATTGGTAATAAAGACTTTGAGCTTGATGCT 2018
QY 708 uGluGluAlaTyrThrThrGluHisTrpLeuValArgIleTyrArgValLys 725
DB 2019 GGAGGAAGCGTATACACAGACACATCGCTGCTAGTCAGGTATACAGGTAAAG 2070

RESULT 12

CNS0ADYZ

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

CNS0ADYZ 2419 bp mRNA linear HTC 06-FEB-2004
Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSUTPGH51ZC09 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
BX816490
BX816490.1 GI:42474670
HTC; GSUT cDNA.
Arabidopsis thaliana (thale cress)

QY 442 AlalSerAlaAlaThrGluValAspGluAlaGluAspSerIleGluLysThrLeu 461
 Db 1452 ACTATCAAGATTAACTTCTCTGTTAAGGACAAACAAAGGTTTCTCACACTGGTTCC 1511
 QY 462 TyrAspLysAlaGlyLysLeuLysHisArgThrLysHisAspAlaGlnAspThrGly 481
 Db 1512 ACGAAAGGAGCTGGTAGTTCAAAGCTTCTTCAAGGTTTACACTTGCATCTCAGCCT 1571
 QY 482 ValSerSerAsnLeuLysSerIleValIleLeuAlaValLeuMetLeuMetMetPhe 501
 Db 1572 TTCAGAGAAATGGTGCACTCTCTCTGTTGTGTATTTTATTTCTCAGTAGATAT 1631
 QY 502 AlavalHisCysThrTyrValThrSerAsnAlaTyrSerProSerIleValLeuAla 521
 Db 1632 GCTATTCTACATGCGATGGTGACAGCAGGCATATTCATCTCCTCAATTGTCTAGCT 1691
 QY 522 PheHisAsnSerGlnAspGlySerArgAsnIleLeuAspAspPheArgGluAlaTyrTyr 541
 Db 1692 GCNAGAGAGGCCAC--GGGAACAGAAATCATCTTTGATGATTACCGGAGGCATATC 1748
 QY 542 TrpLeuSerGlnAsnThrAlaAspAlaArgValMetSerTrpTyrAspTyrGlyTyr 561
 Db 1749 TGGCTTAGGCAAAACACATGCGACCTGATGTAAGATTATGCTGTGGACTATGGGTAC 1808
 QY 562 GlnIleAlaGlyMetAlaAsnArgThrTrpLeuValAspAsnAsnThrTrpAsnAsnSer 581
 Db 1809 CAATCTGCTGCGTGGGAATAGAACTGTCATGCTGCGATAACAATCTCGAACAACACT 1868
 QY 582 HisIleAlaLeuValGlyLysAlaMetSerSerThrGluGluLysSerTyrGluIleMet 601
 Db 1869 CACATTGCTACCGTTGGAGCCGCCATGTCCTTCTATGAGATGATGCTGATGATCATG 1928
 QY 602 ThrSerLeuAspValAspTyrValLeuValIlePheGlyGlyValIleGlyTyrSerGly 621
 Db 1929 AGGTCACCTTGATGTGAATATGATGCTGCTCTTTGCTGCTGTTACTGCTTATCTTCA 1988
 QY 622 AspAspIleAsnLysPheLeuTrpMetValArgIleAlaGluGlyGluHisProLysAsp 641
 Db 1989 GATGATATCAACAGATCTTGTGGATGGTGAGGATGGAGGTGGAGTATTCACAGTG-- 2045
 QY 642 IleLysGluSerAspTyrPheThrAspArgGlyGluPheArgValAspAlaGluGlyAla 661
 Db 2046 ATCAAGAAACCTGATTACCTTGTGAAT--GGTGAATCTCGGTAGACAAAGGTGCATCA 2102
 QY 662 ProAlaLeuLeuAsnCysLeuMetTyrLysLeuSerTyrTyrArgPheGlyGluLeuLys 681
 Db 2103 CCGAAGATGCTGAATCTGCTCATGTACAAAGTTATGCTTATACAGGTTCGGTAGCTGACC 2162
 QY 682 LeuAspTyrArgGlyProSerGlyTyrAspArgThrArgAsnAlaValIleGlyAsnLys 701
 Db 2163 ACAGATACGCAAGCCACCGGATATGATAGACAGAGAGGTGGAGATAGGACAA 2222
 QY 702 AspPheAspLeuThrTyrLeuGluGluAlaTyrThrThrGluHisTrpLeuValArgIle 721
 Db 2223 GATATCAAACTTGAACTTAGAAGAGCTTACACAACTCAACATCCAACTGGATAGTCTGATT 2282
 QY 722 TyrArgValLysLysPro 727
 Db 2283 TACAGAGTTAAACCTCT 2300
 RESULT 13
 LOCUS AY418285 2079 bp DNA linear GSS 17-DEC-2003
 DEFINITION Pan troglodytes ITM1 gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY418285
 VERSION AY418285.1 GI:39774245
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE 1 (bases 1 to 2079)

AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
 Adams, M.D. and Cargill, M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 2079)
 AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sainsky, J.J.,
 Adams, M.D. and Cargill, M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.

FEATURES

Location/Qualifiers
 source
 1..2079
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>2079
 /gene="ITM1"
 /locus_tag="HCM6506"

gene

ORIGIN

Alignment Scores:

Pred. No.: 5,466-178 Length: 2079
 Score: 1669.50 Matches: 332
 Percent Similarity: 60.57% Conservative: 75
 Best Local Similarity: 49.40% Mismatches: 234
 Query Match: 41.26% Indels: 31
 DB: Gaps: 4

US-10-028-384-8 (1-774) x AY418285 (1-2079)

QY 53 PheAsnTyrArgAlaThrAlaTyrMetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrp 72
 Db 154 TTTAATTATCGACTACCAAGNNCTGCTGNNNNNGGGTTTATAAATTCATACTGG 213
 QY 73 PheAspGluArgAlaTyrTyrProLeuGlyArgIleValGlyGlyThrValTyrProGly 92
 Db 214 TTGTGATCCGAGCGCTGTACCTTTGGAGCAATCATTTGGAGGAGCANNNTTACCAGGT 273
 QY 93 LeuMetIleThrSerGlyGlyIleHisTrpLeuLeuHisValLeuAsnIleProValHis 112
 Db 274 TTAATGATCCTCTGCTGCAATCTACCATGTACTCCATTTTTCACATCACCATCGAC 333
 QY 113 IleArgAspIleCysValPheLeuAlaProIlePheSerGlyLeuThrSerIleSerThr 132
 Db 334 ATTCGGAATGCTNTGTCTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAG 393
 QY 133 TyrLeuLeuThrLysGluLeuTrpSerAlaGlyAlaGlyLeuPheAlaAlaSerPheIle 152
 Db 394 TACCACCTTCAAAAGAGCTCAAGATCGAGGGCTGGGCTCTGTGCTGCGCATGATT 453
 QY 153 AlaIleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIle 172
 Db 454 GCTGTAGTTCCTGGATATATCTCCGATCTGCTGGCTCTCTCTCTCTCTCTCTCTCTCT 513
 QY 173 AlalIlePheAlaLeuGlnPheThrTyrPheLeuTrpValArgSerValLysThrGlySer 192
 Db 514 GCCATCTTTTGCATGCTACTCACCTACTACTGATGATGATGATGATGATGATGATGATG 573
 QY 193 ValPheTrpSerAlaAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTrpGlyGly 212
 Db 574 ATCTGTTGGGAGCTAAGTGTGCCCTTGTCTTATTTTACATGCTCTGCTCATGGAGGT 633
 QY 213 TyrValPheIleIleAsnLeuIleProLeuHisValPheValLeuLeuIleMetGlyArg 232
 Db 634 TATGTGTCTCTGATCAACTTAATTCCTCTCCACGCTCTCTCTCTCTCTCTCTCTCTCTCT 693

593 ThrGluGlySerTyrGluLeuMetThrSerLeuAspValAspTyrValLeuValile 612
 1684 ACAGAGAAAGACCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGTCATT 1743
 613 PheGlyGlyValileGlyTyrSerGlyAspAspIleAsnLysPheLeuTyrMetValArg 632
 1744 TTGGAGGCTCCTCGGTTATCTCTGATGATATCAACAAGTTCTNNNGATGTCGG 1803
 633 Ile---AlaGluGlyLysHisProLysAspIleLysGluSerAspTyrPheThrAspArg 651
 1804 ATTGGAGGAGACAGANNANGCAACATATCAGGAGATGACTATATCTCACT 1863
 652 GlyGluPheArgValAspAlaGluAlaProAlaLeuLeuAsnCysLeuMetTyrLys 671
 1864 GGGAGTTCCGTTGNNCNTGAGTTCTCCAGTGTCTCANNNNNNNNNNNNNN 1923
 672 LeuSerTyrTyrArgPheGlyGlyLeuLysLeuAspTyrArgGlyProSerGlyTyrAsp 691
 1924 NNN 1983
 692 ArgThrArgAsnAlaValileGlyAsnLysAspPheAspLeuThrTyrLeuGluAla 711
 1984 CGTCTCGAAATGCTGAGATTGGATTAAGACTTTGACCTTGATGCTCCTGGAGGA 2043
 712 TyrThrThrGluHisTyrLeuValArgIleTyrArg 723
 2044 TATACACAGAACATTGGCTGCTCAGATATACAAG 2079

RESULT 14
 CNS017ZX 836 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
 DEFINITION BACN37M18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL108711.1 GI:5629015
 VERSION AL108711.1
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 836)
 REFERENCE Genoscope.
 AUTHORS Direct Submission
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
 1..836 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACN37M18"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : T7"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.66e-147 Length: 836
 Score: 1396.00 Matches: 274
 Percent Similarity: 98.92% Conservative: 0
 Best Local Similarity: 98.92% Mismatches: 2
 Query Match: 34.50% Indels: 1

233 TyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSer 252
 694 TTCTCTCACCAGATCATGTGGCTACTGTACTGTTTACTGCTGGGCACATACTTTCT 753
 253 MetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaAlaLeu 272
 754 ATGCAGATCTCTTTGTGGGTTTCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 813
 273 GlyValPheValLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSer 292
 814 NNN 873
 293 ArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuValGlyValGlyValPhe 312
 874 CCANAACAATTNNNGTTCTTTCCGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 933
 313 ValAlaValValLeuLeuMetThrLeuGlyValValAlaProTyrSerGlyArgPheTyr 332
 934 NNN 993
 333 SerLeuTyrAspThrGlyTyrAlaLysIleHisIleProIleAlaSerValSerGlu 352
 994 TCACCTGCTGCTCTTCTTATGCTAAGAACACATCCCATCATTTGCTGCTGCTGAG 1053
 353 HisGlnProThrThrTyrPheSerPhePheAspLeuHisIleLeuValCysAlaPhe 372
 1054 CATCAGCCACACACCTGGCTCTACTATTITGACCTNNAGCTCTGCTCTCATGTTT 1113
 373 ProValGlyValTyrTyrCysIleLysGlnIleAsnAspGluArgValPheValLeu 392
 1114 CCAATTGGCTCTATTACTGCTTTNGCAACCTGTCTGATGCCGATTTTATCATCATG 1173
 393 TyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuThr 412
 1174 TATGGTGTGACGACGATCTTTTTCAGCTGTAATG----- 1209
 413 ProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAspValPheLeuGln 432
 1210 -----NNNNNN 1215
 433 GluAspSerSerLysArgMetGlyThrAlaIleSerAlaAlaThrGluValAspGluAla 452
 1216 NNN 1275
 453 GluAspSerIleGluLysLysThrLeuTyrAspLysAlaGlyLysLeuLysHisArgThr 472
 1276 NTGTCNNNATACATGAAGAAANTGNACATAANNNNNNNNNNNNNNNNNNNN 1326
 473 LysHisAspAlaGlnGlnAspThrGlyValSerSerAsnLeuLysSerIleValleLeu 492
 1327 AAGAAGNNNNAGATTCCCCNCCCTATTAAAGATGAAGTGGCAAGTGGGATGATCTG 1386
 493 AlaValLeuMetLeuLeuMetPheAlaValHisCysThrTyrValThrSerAlaAla 512
 1387 GTACGTGCTTTCTTTCTATCACCCTACCTCTTATCACCCTGGGTGACGAGTGGGC 1446
 513 TyrSerSerProSerIleValLeuAlaPheHisAsnSerGlnAspGlySerArgAsnIle 532
 1447 TACTTTCTCGTCCATTGTTACTATCTGCGCGTGGTGGG--GATGGCAGTAGGATCAT 1503
 533 LeuAspAspPheArgGluAlaTyrTyrTyrLeuSerGlnAsnThrAlaAspAlaArg 552
 1504 TTTGATGACTCCGAGAGCATATTATGCTGCTGCTCACAATCTCCAGAGGATGGGAG 1563
 553 ValMetSerTyrTyrAspTyrGlyTyrGlnIleAlaGlyMetAlaAsnArgThrLeu 572
 1564 GTCATNNCTGCTGGATTATGCTATACAGCTATGCAACACCGAACAAATTTTA 1623
 573 ValAspAsnAsnThrTyrAsnAsnSerHisIleAlaLeuValGlyLysAlaMetSerSer 592
 1624 GTGGACATACACATGATGAATAATACCATTTTCTCANNNNNNNNNNNNNNNN 1683

```

DB:          9          Gaps:          0
US-10-028-384-8 (1-774) x CNS017ZX (1-836)

QY 153 AlalalleValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyLe 172
Db 3 GCCATCGTGGCCGTCATCATAGTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 62
QY 173 AlalallePheAlaLeuGlnPheThrTyrPheLeuTyrValArgSerValIleThrGlySer 192
Db 63 GCCATATTCGCCCTGCAGTTCACCTACTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGG 122
QY 193 ValPheThrPheAlaAlaAlaLeuSerTyrPheTyrMetValSerAlaTyrGlyGly 212
Db 123 GTGTTCTGGTGGCCGCGAGCGCTTTGTCTACTTCTACATGGTGTGGCTGGGKGCC 182
QY 213 TyrValPheIleIleAsnLeuIlePheLeuHisValPheValLeuLeuIleMetGlyArg 232
Db 193 TACGTGTTTCATCATCACTCATACCTCCCTGCAGCTTCCTGCTGCTGCTGCTGCTGCT 242
QY 233 TyrSerProArgLeuLeuThrSerTyrSerThrPheTyrIleLeuGlyLeuLeuPheSer 252
Db 243 TACTCGCGCGCTGCTGACGAGCTACAGCACCTTCTACATCTCTGGGACTGCTGTTCTCC 302
QY 253 MetGlnIleProPheValGlyPheGlnProIleArgThrSerGluHisMetAlaLeu 272
Db 303 ATGAGATCCCTCTCGTGGGRTCCACCCGATACGCCACGATGACACATGCTGGCTG 362
QY 273 GlyValPheValLeuLeuMetAlaValAlaThrLeuArgHisLeuGlnSerValLeuSer 292
Db 363 GGRGTGTTTGTGCTTATGCGCGTGGCCACCTTGGCCATTTGCGATTCGCTGCTG 422
QY 293 ArgAsnGluPheArgLysLeuPheIleValGlyGlyLeuLeuValGlyValGlyValPhe 312
Db 423 CGCAACAGAGTTCCGGAAGCTTTCATCGTCGSGGRTTGTGTTGGSGTGGTGGCTTT 482
QY 313 ValAlaValValLeuThrMetLeuGlyValValAlaProIlePheSerGlyArgPheTyr 332
Db 483 GTGGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542
QY 333 SerLeuTrpAspThrGlyTyrAlaLys-IleHisIleProIleIleAlaSerValSerGI 352
Db 543 TCGCTGTGGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
QY 352 uHisGlnProThrThrPheSerPhePheAspLeuHisIleLeuValCysAlaph 372
Db 603 GCATCAGCCACCACTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 662
QY 372 eProValGlyValTrpTyrCysIleLysGlnIleAsnAspGluArgValPheValValle 392
Db 663 CCCAGTGGGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
QY 392 uTyrAlaIleSerAlaValTyrPheAlaGlyValMetValArgLeuMetLeuThrLeuTh 412
Db 723 GTAGCCCATCAGTGGCTTTTACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 782
QY 412 rProValValCysMetLeuAlaGlyValAlaPheSerGlyLeuLeuAsp 428
Db 783 GCCGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 831

RESULT 15
BF502026
LOCUS
DEFINITION
Drosophila melanogaster adult testes pOTB7
Ppan0007748 'enzyme' located on: 3R 96B16-96B17: 04/09/2001, mRNA
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

```

REFERENCE
AUTHORS
Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Baxter, E.,
Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V.,
Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N.,
Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunoo, J.,
Paclob, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C.,
Lewis, S.E., Celniker, S. and Rubin, G.M.
BDGP/HMI Ar Drosophila EST Project
Unpublished (2000)
COMMENT
On Dec 6, 2000 this sequence version replaced gi:11585327.
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AB003569: arm:X [20544566,20835979]
estimated-cyto:1954-19F1: 04/09/2001 hit P element 1(3)j2D9:
1(3)j2D9 A0026308 inserted at base 292 5' end of P element inverse
PCR: 02/10/2001
Plate: AT:176 row: D column: 1
High quality sequence stop: 713.
Location/Qualifiers
1..810
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="AT17637"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT:10-AT:120: DHS-alpha. Plates
AT:121-AT:319: DHS-alpha Tona"
/clone_lib="AT Drosophila melanogaster adult testes pOTB7"
/note="Organ: ADULT testes; Vector: pOTB7; Site: 1: EcorI;
Site: 2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
ORIGIN
Alignment Scores:
Pred. No.: 3,93e-140 Length: 810
Score: 1331.00 Matches: 253
Percent Similarity: 99.61% Conservative: 0
Best Local Similarity: 99.61% Mismatches: 1
Query Match: 32.90% Indels: 0
DB: 2 Gaps: 0
US-10-028-384-8 (1-774) x BF502026 (1-810)
QY 1 MetAsnArgThrProLysMetLeuAsnSerLysValAlaGlyTyrSerSerLeuIleThr 20
Db 49 ATGATTCGAGCGCCGAGATGCTGAACAGCAAGTGGCTGGCTACAGCAGCTTAATACC 108
QY 21 PheAlaIleLeuLeuIleAlaTyrLeuAlaGlyPheSerSerArgLeuPheAlaValle 40
Db 109 TTCGCATCTCTGTAATCGCTGGTGGCGGATTTCTCTCGCTCTTTCGCGCTCATC 168
QY 41 ArgPheGluSerIleIleHisGluPheAspProThrPheAsnTyrArgAlaThrAlaTyr 60
Db 169 CGTTTCGAGTGCATTTATCCATGATTTGATCCGCGGTCAACTACCGGGCCACCGCTAC 228
QY 61 MetValGlnAsnGlyTyrTyrAsnPheLeuAsnTrpPheAspGluArgAlaTyrTyrPro 80
Db 229 ATGTTGCAGATGTTGGTACAACTTCTCACTGTTTCGACGAGCGCGCATGGTATCCG 288
QY 81 LeuGlyArgIleValGlyGlyThrValTyrProGlyLeuMetIleThrSerGlyGlyLe 100
Db 289 CTCGGCAGGATTTGGGCGGTACCGTCTATCTCCCGCTGATGATTACGTCGCGGGAATC 348

```

FEATURES

source

us-10-028-384-8.rst

Thu Dec 16 16:25:17 2004

```

QY 101 HisTrpLeuLeuHisValLeuAsnIleProValHisIleArgAspIleCysValPheLeu 120
Db 349 CATGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATTCGGTGTTCCTG 408
QY 121 AlaProIlePheSerGlyLeuThrSerIleSerThrTyrLeuLeuThrLysGluLeuTrp 140
Db 409 GCGCGGANTTCAGTGGGCTGACCTCCATCTCCACCTACTGCTGACCAAGAGCTGTGG 468
QY 141 SerAlaGlyAlaGlyLeuPheAlaAlaSerPheIleAlaIleValProGlyTyrIleSer 160
Db 469 TCGCGGGCGCGGCTCTTCGCGGCCAGCTTCATCGCCATCGTGGCTGGTACATCAGT 528
QY 161 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheAlaLeuGlnPheThr 180
Db 529 AGTTCGGTGGCTGGATCGTACGATACCGAGGCAATGGCATATTCGCCCTGCAGTTCACC 588
QY 181 TyrPheLeuTrpValArgSerValIlyThrGlySerValPheTrpSerAlaAlaAla 200
Db 589 TACTTCCTGCTGGTGGCTCAGTGAAGACTGGATCCGTGTTCTGGTGGCGCGCAGCCGT 648
QY 201 LeuSerTyrPheTyrMetValSerAlaTrpGlyTyrValPheIleIleAsnLeuIle 220
Db 649 TTATCTACTTCTACATGGTGCCGCTGGGTGGCTAGTGTTCATCAACCTGATA 708
QY 221 ProLeuHisValPheValLeuLeuIleMetGlyArgTyrSerProArgLeuLeuThrSer 240
Db 709 CCCCTGCACGTCTTCGTACTGCTCATTATGGGCAGGTACTCGCCGCGTCTGTGACCAGC 768
QY 241 TyrSerThrPheTyrIleLeuGlyLeuLeuPheSerMetGln 254
Db 769 TACAGCACCTTCTACATNCTGGGACTGCTGTTCTCCATGCAG 810

```

Search completed: December 15, 2004, 11:53:25
Job time : 6096.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 23:35:31 ; Search time 887.605 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTRGFLRLSYEKQDTLLKL.....HMLVRIYKVDLNRGLSRT 705

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8161859

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spc01/US10028384/runat_14122004_131648_6892/app_query.fasta_1.3740
-DB=Published Applications NA -QFW=fastap -SUFFIX=Oligo.rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10028384@cgn1.1.2124 -runat_14122004_131648_6892
-NCPUS=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-Fgapop=6 -Fgapext=7 -Ygapop=60 -Ygapext=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 705 | 100.0 | 2472 | 15 | US-10-171-581-112 |
| 2 | 705 | 100.0 | 2472 | 15 | US-10-028-384-112 |
| 3 | 705 | 100.0 | 2472 | 15 | US-10-173-118-742 |
| 4 | 705 | 100.0 | 2472 | 16 | US-10-343-887-742 |
| 5 | 226 | 32.1 | 3046 | 18 | US-10-417-375-95 |
| 6 | 226 | 32.1 | 5404 | 18 | US-10-417-375-99 |
| 7 | 226 | 32.1 | 5827 | 18 | US-10-417-375-97 |
| 8 | 152 | 21.6 | 3093 | 18 | US-10-417-375-92 |
| 9 | 152 | 21.6 | 3094 | 15 | US-10-028-384-92 |
| c 10 | 146 | 20.7 | 440 | 9 | US-09-878-722-59 |
| c 11 | 146 | 20.7 | 440 | 10 | US-09-904-456-59 |
| 12 | 140 | 19.9 | 452 | 9 | US-09-878-178-1456 |
| 13 | 140 | 19.9 | 452 | 13 | US-10-046-935-1456 |
| 14 | 140 | 19.9 | 452 | 14 | US-10-146-502-1456 |
| 15 | 128 | 18.2 | 487 | 10 | US-09-871-161-213 |
| 16 | 128 | 17.9 | 575 | 9 | US-09-525-297-53 |
| 17 | 102 | 14.5 | 487 | 9 | US-09-736-457-1655 |
| 18 | 102 | 14.5 | 487 | 9 | US-09-302-941-1655 |
| 19 | 102 | 14.5 | 487 | 9 | US-09-849-626-1655 |
| 20 | 102 | 14.5 | 487 | 14 | US-10-017-754-1655 |
| 21 | 102 | 14.5 | 487 | 15 | US-10-113-872-1655 |
| 22 | 102 | 14.5 | 487 | 15 | US-10-283-017-1655 |
| 23 | 101 | 14.3 | 307 | 9 | US-09-736-457-375 |
| 24 | 101 | 14.3 | 307 | 9 | US-09-736-457-1271 |
| 25 | 101 | 14.3 | 307 | 9 | US-09-302-941-375 |
| 26 | 101 | 14.3 | 307 | 9 | US-09-302-941-1271 |
| 27 | 101 | 14.3 | 307 | 9 | US-09-849-626-375 |
| 28 | 101 | 14.3 | 307 | 9 | US-09-849-626-1271 |
| 29 | 101 | 14.3 | 307 | 10 | US-09-476-300-375 |
| 30 | 101 | 14.3 | 307 | 14 | US-10-017-754-375 |
| 31 | 101 | 14.3 | 307 | 14 | US-10-017-754-1271 |
| 32 | 101 | 14.3 | 307 | 15 | US-10-113-872-375 |
| 33 | 101 | 14.3 | 307 | 15 | US-10-113-872-1271 |
| 34 | 101 | 14.3 | 307 | 15 | US-10-283-017-375 |
| 35 | 101 | 14.3 | 307 | 15 | US-10-283-017-1271 |
| 36 | 94 | 13.3 | 419 | 9 | US-09-360-352-10293 |
| 37 | 81 | 11.5 | 596 | 11 | US-09-876-143-200 |
| 38 | 81 | 11.5 | 596 | 11 | US-09-876-143-288 |
| 39 | 80 | 11.3 | 476 | 10 | US-09-918-995-22164 |
| 40 | 65 | 9.2 | 307 | 9 | US-09-736-457-588 |
| 41 | 65 | 9.2 | 307 | 9 | US-09-302-941-588 |
| 42 | 65 | 9.2 | 307 | 9 | US-09-849-626-588 |
| 43 | 65 | 9.2 | 307 | 10 | US-09-476-300-588 |
| 44 | 65 | 9.2 | 307 | 14 | US-10-017-754-588 |
| 45 | 65 | 9.2 | 307 | 15 | US-10-113-872-588 |

ALIGNMENTS

RESULT 1
US-10-171-581-112
; Sequence 112, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 112
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L38961

1 DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-112

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-171-581-112 (1-2472)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTGGATTGCGGATTCCTATGAGAGCAGGACACACTTTTGAAGCTT 166
Qy 21 LeuileLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGCAATGGCTGCTGTATTATCCTCTCCACTCGTCTGTTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTGAAAGTGTATCCATGAGTTGATCCGACTTTTAAATTATCGGACTACCAAGTTCTG 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrPyrProLeu 80
Db 287 GCTCAGGAGGGGTTTATAAATCCATAACTGGTGTGATGACCGAGCCTGGTACCCCTTG 346
Qy 81 GlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGAGCAATCATTTGAGGAAACAATTTACCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTCCACATCACCATCGACATTCGGAATGTCTGTGTCTCTGCTGCC 466
Qy 121 ProLeuSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTTCTGTCTGCTGCGCATGATGCTGTAGTTCCTGGATATATCTCCGA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATTAATGAGGATTTGCCATCTTTTGATGCTACTACCTAC 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGATCAAGGCAGTAAGATGGTTCCATCTGTGGCAGCTAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 707 GCTTATTCTACATGGTCTCTGATGGGAGGTTATGTGTTCTGATCAACTTAATCTCT 766
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACAGTCTCTGCTGATGCTCACAGGCGGTTTCTCTCACCGGATCTATGTGGCCTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTACTGTTTACTCCCTGGGTACTATACCTTTCTAGCAGATCTCTCTGTGGGTTTCCAG 886
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CCTGCTCTTTTCATCAGAGCATGGCAGGCTTTGGGGCTTTTGGTCTCTGCCAGATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db 947 GCCTTTGTGGATTACTTGGCAGCAAGTTCGATCCACACNATTTGAAGTCTTTTTCGG 1006

Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGGTAGGCTTTGCTCTTCTACCGTGGGAGCTCTCTCTCATGCTGACA 1066
Qy 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1067 GGAARAATATCTCCCTGACGCGGGCTTTCTACTCACTGCTGGATCCCTCTTATGTAAG 1126
Qy 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1127 AACCAATCCCCATCANTGCTTCTGTGTGAGCATCAGCCCAACCTGCTCTCATAC 1186
Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1187 TATTTTGACCTGACGCTCTCTCTCTTCTATGTTCCAGTTGGCTCTTACTTACTGCTTTAGC 1246
Qy 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
Db 1247 AACCTGTCTGATGCGCGGATTTTATCATCATATGATGCTGTGACCAAGCATGTACTTTCA 1306
Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1307 GCTGTAAATGGTGGCTCTTAATGCTAGTGTGGACCTCTTATGAGCATTTCTCTGGCAAT 1366
Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
Db 1367 GGAGTCTCTCCAGTCTCTGTCACATACATGAAGAATCTGGACATAAGTCCGCCAGACAAG 1426
Qy 441 LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1427 AGAGCAGAGACACACAGATTCACCTACCTATTAAAGATTGAGTGGCAAGTGGGATG 1486
Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
Db 1487 ATACTGGTCATGGCTTTCTTCTCATCACCTACACCTTTCATTCAACCTGGGTGACCAAT 1546
Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1547 GAGGCTACTCTTCTCCGTCATGCTACTATGCGCGTGGTGGGATGSCAGTAGATC 1606
Qy 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluLysAla 520
Db 1607 ATATTGTATGATTCGAGAGACATATTTATGGTTCGTCATATAATCTCCAGAGATGGC 1666
Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGTTCATGTCCTGGTGGATTTATGGCTATCAGATTACAGCTATGGCAACCGAACAAT 1726
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGGCAATAAACACATGGAATAATACCATATTTCTCAGTAGGAGGAGCAATGGCG 1786
Qy 561 SerThrGluGluLysAlaTyrGluMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGTATGCTGCTC 1846
Qy 581 IlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCTCCTGGGTATCTCTGATGATATCAACAGTTCCTTTGGATGGTC 1906
Qy 601 ArgIleGlyLysThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTTGGAGGAGGACAGATACAGGCAACATATCAAGGAGATGATATATATCTCCA 1966
Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGATTCGCGTGGACCGGTGAGGTTCTCAGTGTGCTGCTCAACTGCCCTCATGTAC 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
Db 2027 AAGATGTGTACTACGCTTTGGACAGGTTTACACAGAGCAAGCTCTCTCCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680

Db 1607 ATATTGATGACTCCGAGAGCATATATTGGCTTCGTCAATAACTCCAGAGGATGCG 1666
Qy 521 LysValMetSerTTPtPpTyrGlyThrAlaMetAlaSerThrile 540
Db 1667 AAGGTCATGCTCGTGGGATTATGGCTATCAGATTACAGCTATGGCAACCGACAAATT 1726
Qy 541 LeuValAspAsnAsnThrThrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGACAAATAACATGGAATAATACCATATTTCTCGAGTAGGAGGAGCAATGGCG 1786
Qy 561 SerThrGluGluValAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTACAGTATGTGTGCTC 1846
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGCCCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCTTTGGATGGTC 1906
Qy 601 ArgIleGlyGlySerThrAspThrGlyIleHisIleGlyGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTGGAGGAGCAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCA 1966
Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGTTCCGTGTGGACCGTGAAGTTCTCCAGTGTGCTCAACTGCCTCATGTA 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrAlaLysArgProProGlyPhe 660
Db 2027 AAGATGTGTACTATCCTTTGACAGGTTTACAGAGAGCAAGGCTCTCCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysPheGluLeuAspValLeuGluGlu 680
Db 2087 GACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCTCGAGAA 2146
Qy 681 GlyTyrThrThrGluHisTTPLeuValAspIleTyrLysValLysAspLeuAspAsnArg 700
Db 2147 GGCTATACCAAGAACATTTGGCTGGTCAGGATATACAGGTAAGAGCCTGGATATCGA 2206
Qy 701 GlyLeuSerArgThr 705
Db 2207 GGTGTGTCAGGACA 2221

RESULT 3

US-10-172-118-742
; Sequence 742, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mac, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE OF INVENTION: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 60/380,770
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002219
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-742

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0
US-10-028-384-12 (1-705) x US-10-172-118-742 (1-2472)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTTGGATTGCTTATGAGAGCAGGACACACATTTTGAAGCTT 166
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCAATGGCTGCTGATATATCCTTCTCCACTGCTGTTGTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTGAAAGTGTATCCATGAGTTTGATCCGTACTTTAATATCGGACTACCGGTTCTTG 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspAlaTyrTyrProLeu 80
Db 287 GCTGAGGAGGGGTTTATAAATTCATTAACCTGGTTTGATGACCGAGGCTGGTACCCCTT 346
Qy 81 GlyValIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACGAATCATTTGGAGGAGCAATTTACCCAGGTTTATGATCACCTCTGCTGCAATCTAC 405
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGACATCGGAATGCGATGCTGTGTCTGCTGCC 466
Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTCTTGGCTGCTGCCATGATGCTGTAGTTCCTGATATATCTCCGA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATAATGAGGGATTCCTATCTTTTGCATGCTACTCACCTAC 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGCAGTAAAGACTGGTTCATCTGTGGGAGGAGTAAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 707 GCTTATTCTACATGGTCTCGTCAATGGGAGGTTATGTTCTCTGATCACTAAATTCCT 766
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACGCTCTCGTGTGATGCTCAGCGCCGTTTCTCTCACCAGGATCTATGTGGCCTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTACTGTCTTACTGCTGGGTACTATATCTTCTAGGCAGATCTCTCTTGTGGGTTTCCAG 886
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CTTGCTCTTCTATCAGACACATGGCAGGGTTGGGGTCTTGTGCTCTCTGCGACATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
Db 947 GCCTTTGTGATTAACCTCGCAGCAAGTTGAATCCACAACAATTTGAAGTCTCTTTCCGG 1006
Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGTAGGCTTTGCTCTTCTCACCCTGGGAGCTCTCTCTCATGCTGACA 1066

QY 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1067 GGAATAATATCTCCCTGGAGGGGGCTTTCTACTCACTGCTGGATCCCTCTTATGCTAAG 1126
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1127 AACAAACATCCCATCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186
QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1187 TATTTTGGACCTGACCTCTGCTCTCATGTTTCCAGTGGCTCTATTAAGCTTTAGC 1246
QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
Db 1247 AACCTGTCTGATGCGCCGATTTTATCATCATGATGATGATGATGATGATGATGATGAT 1306
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1307 GCTGTAATGGTGGCTTAATGCTAGTGTGGCACCCTGTTATGAGCATTCCTCTGGCAT 1366
QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
Db 1367 GGAGTCTCCAGTCTCTCCACATACATGAAGATCTGGACATTAAGTCCGCCAGACAAG 1426
QY 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1427 AAGACCAAGACCAACAGATTCACCTACCTATTAAGATTTGAAGTGGCAAGTGGGATG 1486
QY 461 IleLeuValMetAlaPhePheIleIleThrTyrThrPheHisSerThrTrpValThrSer 480
Db 1487 ATACTGGTCATGGCTTTCTTCTCATCACCTACACCTTTTATCAACCTGGGTGACCACT 1546
QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1547 GAGGCTACTCTTCCCGTCCATGTTACTATCTGCCGTGGTGGGATGGCAGTAGGATC 1606
QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
Db 1607 ATATTGTATGATCTCCGAGAGCATATATTGGCTTCTGTCATATATCTCAGAGGATGGC 1666
QY 521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGGTCATGCTGCTGGGATATGCTATCAGATTCAGCTATGCGCAACCCAGCAAT 1726
QY 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGGCAATAACACATGAATAATACCATATTTCTGAGTAGGCGAGCAATGGCG 1786
QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGAAAAGCCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGCTGCTC 1846
QY 581 IlePheGlyGlyValThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906
QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGATTGGAGGAGGACACATACAGGCAACATATCAAGAGAAATGACTATATATCTCCA 1966
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGGATCCGCTGGAGCGGTGAGGTTCTCCAGTGGCTGCTCAACTGCCTCATGTAC 2026
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2027 AGATGTGTACTATCGTTTGGACAGGTTTACACAGAGCAAGCCCTCTCCAGGCTTT 2086
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2087 GACCGTGTCCGAATGCTGAGATGGGAATAAGACCTTGGAGCTTGATGCTCTGGAGGA 2146
QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700

Db 2147 GCCTATACCAAGCAATGGCTGCTCAGGATATACAGGTAAAGACCTGGATATCGA 2206
QY 701 GlyLeuSerArgThr 705
Db 2207 GCGTTGTCAAGGACA 2221
RESULT 4
US-10-342-887-742
; Sequence 742, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 16/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-742
Alignment Scores:
Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-10-028-384-12 (1-705) x US-10-342-887-742 (1-2472)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTTGGGATTCCTATGAGAGCAGGACACACTTTTGAAGCTT 166
QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGCAATGGCTGCTGCTATATCTTCTCCACTCGTCTGTTCTGCTGCTGAGA 226
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTGAAAGTGTATCCATGAGTTGATCCCTACTTTATATCGACTACCAAGTTCTCTG 286
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
Db 287 GCTGAGGAGGGGTTTTATAAATTCCTAATCTGTTTGTATGACCGAGCCTGGTACCCCTTG 346
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACGATCATCTGAGGAGCAATTTACCCAGGTTTATGATCACTCTGCTGCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTAATCTCCATTTTTCACATCACCATCGAATCGGAATCTGTTGTGTTCTGGGCC 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140

1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 136 ATGACTAAGTTGGATTTTGGGATTTCTATGAGAGAGACACATTTTGAAGCTT 195
 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 196 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTGCTGTTGTGCTGCTGAGA 255
 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 256 TTTGAAAGTGTATCCATGAGTTGATCCGTAATTTAATATCGGACTACAGAGTTCTTG 315
 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaIleTyrProLeu 80
 316 GCTGAGGAGGGGTTTATAAATCCATAAATCGTATGATGACCGAGCTGTGTAACCTTTG 375
 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 376 GGACCAATCATTTGGAGGAACATTTACCCAGGTTTATGATCATCCTCTGCTGCAATCTAC 435
 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 436 CATGTACTCCATTTTCCACATCACCATCGACATTCGGAATGCTGTGTGTTCTGCGCC 495
 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA 140
 496 CCTCTCTTCT 553
 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
 554 ATGCAGGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 613
 160 qSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
 614 CATCTGTGGCTGCTCTCTATGATATGAGAGGATTTGCACTTTTGGATCTCTACTCACCT 673
 180 yTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaL 200
 674 ACTACATGTGGATCAAGCAGCAGTAAGACTGGTTCATCTGTGTGGCAGCTAAGTGTGCC 733
 200 euAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuAsnLeuLe 220
 734 TTGCTTATTTCTACATGCTCTGCTCATGGGAGGTTATGTTCTCTGATCACTTATTC 793
 220 roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 240
 794 CTCCTCAGCTCTCTGCTGCTGCTCATCAGGCGCTTTCTCTCAACCGATCTATGTGCGCT 853
 240 yCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 854 ACTGTACTGTTTACTGCTGGCCTATACCTTTCTAT-GCAGATCTCTCTTGTGGGTTTC 912
 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
 913 CAGCTCTGCTCTTCTATCAGAGCAGACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGAT 971
 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPh 299
 972 CCAATGCTTTGTGGATTACCTCGGAGCAAGTTGAATCCACCAATTTGAAGTTCTTTT 1031
 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe 319
 1032 CCGGAGCGTCTCTCTGTAGGCTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1091
 319 uThrGlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTrpAl 339
 1092 GACAGAAATATCTCTCTGACGGGGGCTTTCTACTCGCTGTGGATCCCTCTCTATGC 1151
 339 aLysAsnIleProIleIleAlaSerValSer-GluHisGlnProThrThrTrpSerSe 359
 1152 TAAGAACAAACATCCCATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1211

359 rTyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
 1212 ATACTATTTGACCTGAGCTCTCTGCTTTCATGTTTCCAGTTGGCTCTATTAAGCTTT 1271
 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh 399
 1272 TAGCAACCTGCTGATGCCGGATTTTATCATCATGTATGGGTGACAGCATGTACTT 1331
 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
 1332 TTAGCTGTAATGGTGGCTCTAATGCTAGTGTGGACCTGTTATGTG-CATTTCTCTCTG 1390
 419 lYlleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
 1391 GCATTGGAGTCTCCAGGCTCTGCCACATACATGAAGAAATCTGGACATAAAGTCGCGCAG 1450
 439 spLysLysSerLysLysGlnAspSerThrTyrProIleLysIle-GluValAlaSer 458
 1451 ACAGAGAGCAAGAGCAACAGGATTCACCTACCTTATTAAGAA-TGAAGTGGCAAGT 1509
 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal 478
 1510 GGGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1569
 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
 1570 ACCAGTGGGCTACTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1629
 499 ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
 1630 AGGATCATATTGATGACTTCCGAGAGAGCATATATTGGCTCTGCTCATATTAATCTCAGAG 1689
 519 AspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
 1690 CATCGAAGGTCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1749
 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
 1750 ACAATTTTAGTGGACAATAACACATGGGAATTAATACCATATTTCTCGAGTAGGCGAGCA 1809
 559 MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal 578
 1810 ATGGCTCCACAGAGGAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGTATGTG 1869
 579 LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrp 598
 1870 CTGCTCAATTTTGGAGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1929
 599 MetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyr 618
 1930 ATGCTCCGATTTGAGGAGCAGACAGATACAGCAACATATCAAGGAGATGACTATTAT 1989
 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
 1990 ACTCCAACTGGGAGTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2049
 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
 2050 ATGTACAGATGCTGTACTATCGCTTTGGACAGTTTACACAGAGCCAGGCTCTCTCCA 2109
 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeu 678
 2110 GGGTTGACCTGCTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTGCTG 2169
 679 GluGlu 680
 2170 GAGGAA 2175

RESULT 6

US-10-417-375-99
 ; Sequence 99, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:

; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 99
 ; LENGTH: 5404
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-417-375-99

Alignment Scores:
 Pred. No.: 1,61e-227 Length: 5404
 Score: 226.00 Matches: 674
 Percent Similarity: 98.25% Conservative: 0
 Best Local Similarity: 98.25% Mismatches: 6
 Query Match: 32.06% Indels: 12
 DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-99 (1-5404)

| | | | |
|----|-----|--|-----|
| Qy | 1 | MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu | 20 |
| Db | 136 | ATGACTAAGTTTGGATTTTGGATTGCTCTATGAGAGCAGGACACACTTTTGAAGCTT | 195 |
| Qy | 21 | LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg | 40 |
| Db | 196 | CTCAATCTGCAATGGCTGCTGATTATCTCTCCACCTGCTGTTGCTGCTCTGAGA | 255 |
| Qy | 41 | PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu | 60 |
| Db | 256 | TTTCAGAGTGTATCCATGATGTTGATCCGTCATTTAATATCCGAGTACCAAGTTCCTG | 315 |
| Qy | 61 | AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu | 80 |
| Db | 316 | GCTGAGGAGGGGTTTATAAATTCATACCTGTTTGTATGACGAGCCTGGTACCTTTG | 375 |
| Qy | 81 | GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr | 100 |
| Db | 376 | GGACGAATCATGGAGGACAAATTTACCCAGGTTTAAATGATCACCTCTGCTCAATCTAC | 435 |
| Qy | 101 | HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla | 120 |
| Db | 436 | CATGTACTCCATTTTTCACATACATGACATCGAATTCGGAATGTCGTGTGTCTCTGCC | 495 |
| Qy | 121 | ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA | 140 |
| Db | 496 | CCTCTCTCTCTCTCTCCACAC-CATCGTCACTGATACCA-CCTTACCAAGAGCTCAAGG | 553 |
| Qy | 140 | spAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerA | 160 |
| Db | 554 | ATGACGGGCTGGGCTCTTGTGCTGCCATGATGTTGCTGTAGTCTCTGTGATATATCTCC | 613 |
| Qy | 160 | rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr | 180 |
| Db | 614 | GATCTGTGGCTGGCTCTATGATAATCAAGGATTCGCATCTTTTGCATGCTACTCACT | 673 |
| Qy | 180 | yrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaL | 200 |
| Db | 674 | ACTACATGTGATCAAGGAGTAAAGACTGGTTCATCTGTTGGGAGCAGTAAGTGTGCC | 733 |
| Qy | 200 | euAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuIleP | 220 |
| Db | 734 | TTGCTTATTTCTACATGGCTCTGCTGATGGGAGGTTATGTTGCTCTGATCACTTAATTC | 793 |
| Qy | 220 | roleuHisValLeuValLeuMetIleThrClyArgPheSerHisArgIleTyrValAlaI | 240 |
| Db | 794 | CTCTCCAGCTCCTGCTGATGCTACAGGCGGTTCTCTCACCAGATCATGTGGGCT | 853 |
| Qy | 240 | yrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe | 259 |

| | | | |
|----|------|---|------|
| Db | 854 | ACTGTACTGTTTACTGCTGGGCACTATACCTTCTAT-GCAGATCTCTCTTTGGGGTTTC | 912 |
| Qy | 260 | GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIle | 279 |
| Db | 913 | CAGCCCTGCTCTTTCATCAGAGCAGCATGCGAGC-CTTTGGGGGTCTTTGGTCTTCCAGAT | 971 |
| Qy | 279 | etisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPh | 299 |
| Db | 972 | CCATGCTTTGTGGATTACCTGCGCAGCAAGTTGAATCCACCAATTTGAAGTCTCTTT | 1031 |
| Qy | 299 | eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetIle | 319 |
| Db | 1032 | CCGAGGCTCATCTCTGCTAGGCTTGTCTCTCTACCGTGGGAGCTCTCTCATGCT | 1091 |
| Qy | 319 | uThrGlyLysIleSerProTrpThrGlyArgPheTyr-SerLeuLeuAspProSerTyrAl | 339 |
| Db | 1092 | GACAGGAAAAATATCTCCCTGACGGGGCGTCTTCTACTCGCTGTGGATCCCTCTTATGC | 1151 |
| Qy | 339 | alysAsnIleProIleIleAlaSer-ValSerGluHisGlnProThrThrTrpSerSe | 359 |
| Db | 1152 | TAAGAACACATCCCATCATGCTTCTGTGCTTGAGCATCAGCCCAACCTGGTCTC | 1211 |
| Qy | 359 | rTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGly-LeuTyrTyrCysph | 379 |
| Db | 1212 | ATACTATTTTACCTGACCTCTCGCTCTCATGTTTCCAGTTGGCCTCTATTACTGCTT | 1271 |
| Qy | 379 | eSerLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh | 399 |
| Db | 1272 | TAGCAACCTGCTGATGCCGATTTTATCATGATGATGATGATGATGATGATGATGAT | 1331 |
| Qy | 399 | eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG | 419 |
| Db | 1332 | TTTCAGCTGTAATGCTGCTCTAATGCTAGTGTGGCCTCTTATGCTG-CATTCTCTCTG | 1390 |
| Qy | 419 | lyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA | 439 |
| Db | 1391 | GAATGGAGTCTCCAGGCTGCTCCACATACATGAGATCTGGACATAGTCTGCTCCAG | 1450 |
| Qy | 439 | spLysSerLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSer | 458 |
| Db | 1451 | ACAAGAGAGCAAGAGCAAGAGGATTCACCTACCTCTTATTAAGAA-TGAAGTGGCAAGT | 1509 |
| Qy | 459 | GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpVal | 478 |
| Db | 1510 | GGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 1569 |
| Qy | 479 | ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer | 498 |
| Db | 1570 | ACCAGTGGGCTACTCTCTCCGTCCTTCTACTATCTGCTGCTGCTGCTGCTGCTGCTG | 1629 |
| Qy | 499 | ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu | 518 |
| Db | 1630 | AGATCATATTTGATGATCTTCGAGAGCATATTTATGGCTTCTGCTCATATATCTCCAG | 1689 |
| Qy | 519 | AspAlaLysValMetSer-TrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg | 538 |
| Db | 1690 | GATGCAAGGTCATGCTCTGCTGGGATTTATGGCTATCAGATTCAGCTATGGCAACCGA | 1749 |
| Qy | 539 | ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla | 558 |
| Db | 1750 | ACAATTTAGTGAGCAATACACATGGAATATATACCATATATTTCTCAGGTAGGGCAGCA | 1809 |
| Qy | 559 | MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal | 578 |
| Db | 1810 | ATGGGCTCCACAGAGAAAAAGCCTATGAGATCATGAGGAGCTGATGTGATGTG | 1869 |
| Qy | 579 | LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrp | 598 |
| Db | 1870 | CTGGTCATTTTGGAGGCTCCTGCGGATTTCTCTGATGATCAACAAGTTCTTTGG | 1929 |
| Qy | 599 | MetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluLeuAspTyrTyr | 618 |

Db 1930 ATGGTCGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGAACTGACTATTAT 1989
 Qy 619 ThrProThrGlyGluPheArgValAspArgGluCysSerProValLeuLeuAsnCysLeu 638
 Db 1980 ACTCCAACTGGGAGTTCCTGGTGGACCGTGAAGGTTCTCCAGTGTCTCAACTCCCTC 2049
 Qy 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
 Db 2050 ATGTACAAAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCCAGCGTCTCCA 2109
 Qy 659 GlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysPheGluLeuAspValLeu 678
 Db 2110 GCTTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCCTG 2169
 Qy 679 GluGlu 680
 Db 2170 GAGGAA 2175

RESULT 7

US-10-417-375-97
 ; Sequence 97, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 97
 ; LENGTH: 5827
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-417-375-97

Alignment Scores:

P-Seq. No.: 1,73e-227 Length: 5827
 Score: 226.00 Matches: 674
 Percent Similarity: 98.25% Conservative: 0
 Best Local Similarity: 98.25% Mismatches: 6
 Query Match: 32.06% Indels: 12
 DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-97 (1-5827)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 136 ATGACTAAGTTGGATTTTTCGGATTGCTCTATGAGAGAGAGAGACACACTTTTGAAGCTT 195
 Qy 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 196 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTGCTGTGTGTGTGCTGCTGAGA 255
 Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 Db 256 TTGAAAGTGTATCCATGAGTTTGAATCCGTTACTTTAATATCGGACTACAGGTTCCCTG 315
 Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu 80
 Db 316 GCTGAGAGAGGGGTTTTATAAATTCATAACTGGTTTGTATGACCGAGGCTTGGTACCCTTTG 375
 Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 376 GGACGAAATCATGGAGGAGACAAATTTACCGAGTTTATGATCACTCTGCTGCAATCTAC 435
 Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 436 CATGACTCCATTTTTCACATCACCATCGACATCGGAATGCTGTGTGTTCTTCCTGGCC 495
 Qy 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA 140

Db 496 CCT 553
 Qy 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
 Db 554 ATGACAGGGCTGGGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 613
 Qy 160 rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
 Db 614 GATCTGTGGCTGGCTCTCTATGATAATGAAGGAGTTCCTATCTCTCTCTCTCTCTCTCTCT 673
 Qy 180 yrTyrMetTyrIleLysAlaValLysThrCysIleCysTrpAlaAlaLysCysAla 200
 Db 674 ACTACAGTGATCAAGGAGTAAAGAGCTGTTCCATCTGTTGGGCGAGTAAAGTGGTCC 733
 Qy 200 euAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuLeuLeuP 220
 Db 734 TTGCTTATTTCTACATGCTCTCTGATGGGAGGTTATGTGTTCTCTGATCAACTTAATTC 793
 Qy 220 roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 240
 Db 794 CTCTCCACGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 853
 Qy 240 yrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 Db 854 ACTGCTACTGTTTACTGCTGGGCACTATCTTTCTAT-GCAGATCTCTCTTTGTGGTTC 912
 Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLysCysGln 279
 Db 913 CAGCTGTCTCTTCTCATCAGAGCACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGAT 971
 Qy 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnInPheGluValLeuP 299
 Db 972 CCATGCTTTGTGGATTACCTGGCGCAGCAAGTGAATCCACACAATTTGAAGTCTCTTT 1031
 Qy 299 eArgSerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuMetIle 319
 Db 1032 CCGAGCGCTCATCTCTGTAGGCTTTGTCTCTCTCACCGTGGAGCTCTCTCTCACT 1091
 Qy 319 uThrGlyLysSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
 Db 1092 GACAGGAAAAATATCTCTCTGGAGGGCGTTTCTACTCGTCTGGATCCCTCTTATGC 1151
 Qy 339 aLysAsnAsnIleProIleAlaSerValSerGluHisGlnProThrTrpSerSe 359
 Db 1152 TAAGAACAAATCCCCATCATCTCTGTGTCTGAGCATCAGCCACCAACCTGGCTCTC 1211
 Qy 359 rTyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysP 379
 Db 1212 ATACTATTTGACCTGCGAGCTCTCTGCTCTCTATGTTTCCAGTTGGCTCTTATCTGCT 1271
 Qy 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrP 399
 Db 1272 TAGCAACCTGCTGATCCCGGATTTTATCATCATGTATGGTGTGACGAGCATGTACTT 1331
 Qy 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
 Db 1332 TTCAGCTGTAATGTGTGCTAATGCTAGTGTGGCACCCTGTTATGTG-CATTCTCTCTG 1390
 Qy 419 lYleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
 Db 1391 GCATGTGAGTCTCCAGGTGCTGTCCACATACATGAAGATCTGGACATGAAGTGGCCAG 1450
 Qy 439 spLysLysSerLysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSer 458
 Db 1451 ACAAGAGAGCAAGAGCAACAGGATTCACCTACCTATTAAGAA-TGAAGTGGCAAGT 1509
 Qy 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal 478
 Db 1510 GGGATGATACATGGTATGGCTTTCTCTCATCACTACACCTTTTATTCAACCTGGGTG 1569
 Qy 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
 Db 1570 ACCAGTGGGCTACTCTTCTCGCTCCATGTACTATCTGCTGGTGGGATGGCAGT 1629

```
OY 499 ArgillePheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGlu 518
Db 1630 AGGATCATATTGATGACTCCGAGAGCATATATTGGCTCGTCATATACTCCAGAG 1689
OY 519 AspAlaValMetSerTyrTyrAspTyrGlyGlyGlyGlyGlyGlyGlyGlyGly 538
Db 1690 GATCGGAGGTCATGCTCTGGGATTATGGCTATCAGATACAGTATGCAACCGA 1749
OY 539 ThrLeuValAspAsnAsnThrTyrAsnAsnThrHisLeuSerArgValGlyGlnAla 558
Db 1750 ACAATTTAGTGACAAATAACACATGGAATAATACCCATATTCTCGAGTAGGCGCA 1809
OY 559 MetAlaSerThrGluGluLysAlaTyrGluLeuMetArgGluLeuAspValSerTyrVal 578
Db 1810 ATGGCTCCACAGAGGAAAAGGCTATGATCATGAGGAGCTCGATGTCAGCTATGTG 1869
OY 579 LeuValLeuPheGlyGlyLeuThrTyrSerSerAspAspLeuAsnLysPheLeuTyr 598
Db 1870 CTGGTCATTTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAACAAAGTTTCTTTGG 1929
OY 599 MetValArgGlyGlySerThrAspThrGlyLysHisLeuSerGluAsnAspTyrTyr 618
Db 1930 ATGGCTCGGATGGAGGAGACACAGATACAGGCAACATATCAAGAGATGACATATAT 1989
OY 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
Db 1990 ACTCCAACCTGGGAGTTCCGTGTGGACCGTGAAGGTTCCTCCAGTGTGCTCAACTGCTC 2049
OY 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPro 658
Db 2050 ATGACAGATGCTGTTACTATCTCTTGGACAGTTTACAGAGGCTTACAGAGGCTCTCCA 2109
OY 659 GlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGluLeuAspValLeu 678
Db 2110 GGCTTTGACCGTGTCCGAAATGCTGAGATGGGAATAAAGACTTTGAGCTGATGCTGCTG 2169
OY 679 GluGlu 680
Db 2170 GAGGAA 2175

RESULT 8
US-10-417-375-92
; Sequence 92, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 3093
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-417-375-92

Alignment Scores:
Pred. No.: 1,36e-149 Length: 3093
Score: 152.00 Matches: 667
Percent Similarity: 97.37% Conservative: 0
Best Local Similarity: 97.37% Mismatches: 9
Query Match: 21.56% Indels: 18
DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-92 (1-3093)

OY 5 GlyPheLeuArgTyrSerTyrGluLysGlnAspThrLeuLeuLysLeuLeuSer 24
Db 124 GGATTTTGGATTGCTTCTATGAGAAGAGGACACACTTCTAAAGCTTCTCATCTGCTG 183
```

```
OY 25 MetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerVal 44
Db 184 ATGGCTGCTGCTGTTATCTTTTCTACTCGTCTTTTCTGCTGCTGAGATTTGAAAGTGTC 243
OY 45 IleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluGly 64
Db 244 ATCCATGAGTTTGTATCCGTAATTTTAAATATCGGACTACCCCGTTTCTGGCTGAGAGGG 303
OY 65 PheTyrLysPheHisAsnTyrPheAspArgAlaTyrTyrProLeuGlyArgIleIle 84
Db 304 TTTTATAATTCCATAAATCTGGTTTGTATGACCGGCTTGGTACCTTTGGGCGCAATCAT 363
OY 85 GlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHis 104
Db 364 GAGGAGACAAATTTCCAGGTTTAAATGATCACTTCTGCTGCATCTACCAATGACTCCAT 423
OY 105 PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSer 124
Db 424 TTTCTTCCATATCACATATTGACATTCGGAATGCTGCTGTTTCTCTGGGCCCACTTTCTCC 483
OY 125 SerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysAspAlaGluVal 144
Db 484 TCTTTCACACAC-CATCGTTAGTACCA-CCTTACCAAGAGCTCAAGATGCGAGAGCTG 541
OY 144 IleLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAlaG 164
Db 542 GGCTTCTGCTGCTGCCATGATGCTGATGTTCTCTGGGTATATTTCTCGATCTGTAGCTG 601
OY 164 IleSerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrp 184
Db 602 GTCTCTGATGATGAAAGAAATTTGCTATCTTTGCTGCTGCTTACTTACTTACTCATGTGGA 661
OY 184 IleValAlaValLysThrGlySerIleCys-TrpAlaAlaLysCysAlaLeuAlaTyrPhe 203
Db 662 TCAAGCGAGTGAAGACTGGTTCATCTA-TTGGGCTGCCAAGTGTGCCCTGCTTATTTC 720
OY 204 TyrMetValSerSerTrpGlyGlyTyrValPheLeuIleLeuLeuLeuProLeuHisVal 223
Db 721 TTACATGGTCTCTTCATGGGAGGCTATGTTCTCTGATCACTTGTATCTCTACATGTC 780
OY 224 LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243
Db 781 CTGGTGTCTAATGCTACAGGCGCTTTTCTCACCGATCTACGTAGCTACTGTACTGTT 840
OY 244 TyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGlnProValLe 263
Db 841 TACTGCTGGGACCATCTCTTCTAT-CCAGATTCTCTTTGTTGTTTCCAGCCCTCTCT 899
OY 263 uSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIleHisAlaPheV 283
Db 900 TTTCTACAGAACACATGGCAGC-CTTTGGAGTGTTTGGTCTCTGTCTAGATCCATGCTTCC 958
OY 283 AlaSerTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArgSerValI 303
Db 959 TGAATTACTGCGCAGCAAGTTGAATCCACAGCAATTCGAAGTTCTTTTCCGAGAGTTA 1018
OY 303 leSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysI 323
Db 1019 TCTCCCTGTTGGCTTGTCTCTCTCACTGTGGAGCTCTCTCATCTAACAGGAAAAA 1078
OY 323 leSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsnI 343
Db 1079 TTTCTCCCTGGACAGGCGTTTCTCTCTCTGCTGGATTCCTCTATGCTAAGAAATACA 1138
OY 343 leProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrTyrPheA 363
Db 1139 TTTCCCATTTATGATCTCTGTTCTGAGCACCAGCCACAACTGCTCTCTCTTCTTATTTG 1198
OY 363 sLeuGlnLeuLeuValPheValPheProValGlyLeuTyrTyrCysPheSerAsnLeuS 383
Db 1199 ATCTACAGCTCTCTGCTCTCTCAATTTTCAAGTTGGCTCTTATTACTGCTTTAGCACTGT 1258
```



```

QY 383 exAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSerAlaValM 403
Db 1259 CTGATGCTCGGATTTTATCATCATGATGATGATGATGATGATGATGATGATG 1318
QY 403 etValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerGlyIleGlyVal 422
Db 1319 TGGTGGCTTAATGCTGATATGGACCTGTTATGTG-CATTCTTTCTGCAATGGTGT 1377
QY 423 SerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSer 442
Db 1378 TCCAGGTCCTGCTCCACATATATGAANAATCTGGACATAAGTCCGCCACAGAGAGC 1437
QY 443 LysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSerGlyMetIleLe 462
Db 1438 AAGAAGCAACAGGATCTTACTTACCTATTAAGAA-TGAGGTGGCGAGTGGGATGACT 1496
QY 462 uValMetAlaPhePheLeuIleThrTyrPheHisSerThrTyrValThrSerGluAl 482
Db 1497 GGTCAATGGCTTTTCTCATCACTACACCTTCACTTGGGTGGGTGACAGTGAAGC 1556
QY 482 atYrSerSerProSerIleValLeuSerAlaArgGlyClyAspGlySerArgIleLePhe 502
Db 1557 CTATTTCTTCTCCTCCATTTACTGCTGCTGGTGGGATGGCAGTAGGATCAATTTT 1616
QY 502 eAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLysVa 522
Db 1617 TGATGACATCCGAGAGAGGATATTTGCTGCTCCGTCACAACTCCAGAGGATGCAAAAGT 1676
QY 522 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVa 542
Db 1677 CATGTCATGGTGGGATTTATGCTACCAATTTACTGCAATGGCAATCGGCAATTTTAGT 1736
QY 542 tAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerTh 562
Db 1737 GGACATTAACATGGATATATACCATATTTCTCGAGTAGGGAGGCAATGGCATCCAC 1796
QY 562 rGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 1797 AGAAGAAAAGCCTATGAATCATGAGGAGCTTGATGTCAGCTATGCTTGTCTCATTTT 1856
QY 582 eGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIle 602
Db 1857 TGGAGGCTTACTGGTATTTCTTGGATGATATCACAGTTCTTTGGATGTCGGAT 1916
QY 602 eGlyGlySerThrAsp-ThrGlyLys-HisIleLysGluAsnAspTyrTyrTrpThr 621
Db 1917 TGGAGGAAGCACAGA-GACAGG-AAGACACATTAAGGAGATGACTACTATCTCTACT 1974
QY 622 GlyGluPheArgValAspGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
Db 1975 GGGGAATCCGTTGATCGTAGGGTTCTCCGTGCTGCTCAACTGCCCTATGTACAAA 2034
QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
Db 2035 ATGTGTACTACCGCTTTGGCGAGTCTACACAGAGCAAGCGTCCACAGGCTTTGAC 2094
QY 662 ArgValArgAsnAlaGluIleGlyValLysPheGluLeuAspValLeuGluGlu 680
Db 2095 CGTGTTCGAATGCTGAGATTTGTAATAAGACTTTGAGCTTGATGCTCGGAGAA 2151

```

RESULT 9

```

US-10-028-384-9
; Sequence 9, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028.384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13

```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_008408
; DATABASE ENTRY DATE: 2000-11-01
; RELEVANT RESIDUES: (1)..(3094)
US-10-028-384-9

Alignment Scores:
Pred. No.: 1,366-149 Length: 3094
Score: 152.00 Matches: 667
Best Similarity: 97.37% Conservative: 0
Best Local Similarity: 97.37% Mismatches: 9
Query Match: 21.56% Indels: 18
DB: 15 Gaps: 0

```

US-10-028-384-12 (1-705) x US-10-028-384-9 (1-3094)

```

QY 5 GlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeuLeuLeuSer 24
Db 124 GGATTTTGGCATTTGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 183
QY 25 MetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerVal 44
Db 184 ATGGCTGCTGCTGATCTTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 243
QY 45 IleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGly 64
Db 244 ATCCATGAGTTTATCGGACTACCCGGTTTCTGGCTGAGGAGGGG 303
QY 65 PheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeuGlyArgIleIle 84
Db 304 TTTTATAAATTCCTACTGTTTGGTACCGGCTTGGTACCCCTTTGGCCCAATTCAT 363
QY 85 GlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHis 104
Db 364 GGAGGAACAAATTTACCAGAGTTTAAATGATCATCTCTGCTGCAATACCATGATCAT 423
QY 105 PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSer 124
Db 424 TCTTTCCATATCATCTATGCAATTCGGAATGCTGTTTCTGGCCCACTTTTCTCC 483
QY 125 SerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysAspAlaGlyAlaG 144
Db 484 TCTTTCCACCAC-CATGCTTACGTACCA-CCTTACCAAGAGAGCTCAAGGATGAGGAGCTG 541
QY 144 LysLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArgSerValAlaG 164
Db 542 GGCTTTTGGCTGCTGCAATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 601
QY 164 LysTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrpI 184
Db 602 GCTCTCATGATAAATGAAGGAATTTGCTATCTTTTGCATGCTGCTTACTTACTACATGGA 661
QY 184 LeLysAlaValLysThrGlySerIleCys-TrpAlaAlaLysCysAlaLeuAlaTyrPhe 203
Db 662 TCAAGGCACTGAAGAGCTGGTTTCCATCTA-TTGGGCTGCCAAGTGTGCCCTCGCTTATTC 720
QY 204 TyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIleProLeuHisVal 223
Db 721 TACATGCTCTCTTCAATGGGAGGCTATGCTTCTGATCACTTGAATCTCTTACATGTC 780
QY 224 LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243
Db 781 CTGTGCTTAATGCTGAGAGCGCGTTTTTCTCACCGGATCTACGCTAGCTACTACTGTGT 840
QY 244 TyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGlnProValle 263
Db 841 TACTGCTGGGACCACTTCTCTAT-GCAGATTTCTTTGTTGTTTCCAGGCCCTCTCT 899

```



```
Db 19 CTGCTCAACTGCCTCATG 2
RESULT 11
US-09-904-456-59/c
; Sequence 59, Application US/09904456
; Publication No. US20030017167A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.524C1
; CURRENT APPLICATION NUMBER: US/09/904,456
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 247
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-456-59
Alignment Scores:
Pred. No.: 4,566-144 Length: 440
Score: 146.00 Matches: 146
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.71% Indels: 0
DB: 10 Gaps: 0
US-10-028-384-12 (1-705) x US-09-904-456-59 (1-440)
QY 494 GlyGlyAspGlySerArgIlePheAspAspPheArgGluAlaTyrTyrTyrLeuArg 513
Db 439 GGTGGGATGGCAGTAGGATCATATTTGATGACTCCGAGAAGCATATATTGGCTTCGT 380
QY 514 HisAsnThrProGluAspAlaValMetSerTyrTyrTyrTyrGlyTyrGlnIleThr 533
Db 379 CATATATCTCCAGAGATCCGAGGTCTATGTCCTGGTGGGATATGCTATCATGATTACA 320
QY 534 AlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTyrAsnAsnThrHisLeuSer 553
Db 319 GCTATGGCAACCCGAACAAATTTAGTGGACAAATTAACACATGGAATATATCCCATATTTCT 260
QY 554 ArgValGlyGlnAlaMetAlaSerThrCluGluLysAlaTyrGluIleMetArgGluLeu 573
Db 259 CGAGTAGGCGAGCAGTGGCTCCACAGAGAAAAGCCTATGAGATCATGAGGGAGCTC 200
QY 574 AspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIle 593
Db 199 GATGTCAGCTATGCTGCTCATTTTGGAGGCTCCTGCTGATGATC 140
QY 594 AsnLysPheLeuTyrMetValArgIleGlySerThrAspThrGlyHisLeuLys 613
Db 139 AACAAATTTCTTTGGATGGTCCGATTGGAGGAGCAGATACAGCAACATATCAAG 80
QY 614 GluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProVal 633
Db 79 GAGAATGACTATTATATCTCAACTGGGAGTTCCGTGGACCGTGGAGGTTCTCCAGTG 20
QY 634 LeuLeuAsnCysLeuMet 639
Db 19 CTGCTCAACTGCCTCATG 2
RESULT 12
US-09-878-178-1456
; Sequence 1456, Application US/09878178
; Patent No. US2003017552A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1456
; LENGTH: 452
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1456
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc feature
; LOCATION: (1)-(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1456
Alignment Scores:
Pred. No.: 1,02e-137 Length: 452
Score: 140.00 Matches: 140
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.86% Indels: 0
DB: 9 Gaps: 0
US-10-028-384-12 (1-705) x US-09-878-178-1456 (1-452)
QY 490 LeuSerAlaArgGlyGlyAspGlySerArgIlePheAspAspPheArgGluAlaTyr 509
Db 2 CTATCTGCCCTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCGAGAAGCATAT 61
QY 510 TyrTyrLeuArgHisAsnThrProGluAspAlaValMetSerTyrTyrTyrTyrGly 529
Db 62 TATTGGCTTCGTATAATCTCCAGAGGATGCCAAGGTATGCTCTGGTGGGATATGCG 121
QY 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTyrAsnAsn 549
Db 122 TATCAGATTACAGCTATGGCAACCCGAACAAATTTAGTGGACAAATTAACACATGGAATAAT 181
QY 550 ThrHisLeuSerArgValGlyGlnAlaMetAlaSerThrCluGluLysAlaTyrGluIle 569
Db 182 ACCATATTTCTCCAGTAGGGCAGGCAATGGCTCCACAGAGAAAAGCCTATGAGATC 241
QY 570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSer 589
Db 242 ATGAGGGAGCTCCGATGTCAGTATGCTGCTCATTTTGGAGGCTCCTCACTGGGTATTC 301
QY 590 SerAspAspIleAsnLysPheLeuTyrMetValArgIleGlySerThrAspThrGly 609
Db 302 TCTGATGATATCAACAAGTTTCTTTGGATGGTCCGATTTGGAGGAGCAGATACAGGC 361
QY 610 LysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGlu 629
Db 362 AACATATCAAGGAGATGACTATTATCTCAACTGGGAGTTCCGTGTGGACCGTGAA 421
RESULT 13
US-10-046-935-1456
; Sequence 1456, Application US/10046935
; Publication No. US20020156011A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; APPLICANT: Wang, Aljun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527C1
; CURRENT APPLICATION NUMBER: US/10/046,935
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2239
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1456
; LENGTH: 452
```

TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 422
 OTHER INFORMATION: n = A,T,C or G
 US-10-046-935-1456

Alignment Scores:
 Pred. No.: 1,028-137 Length: 452
 Score: 140.00 Matches: 140
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.86% Indels: 0
 DB: 13 Gaps: 0

US-10-028-384-12 (1-705) x US-10-046-935-1456 (1-452)

QY 490 LeuSerAlaArgGlyGlyAspGlySerArgGlyIlePheAspAspPheArgGluAlaTyr 509
 DB 2 CTATCTCCCGTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCCGAGAAGCATAT 61
 QY 510 TyrTrpLeuArgHisAsnThrProGluAspAlaValMetSerTrpAspTyrGly 529
 DB 62 TATTGGCTTCGTATATATCTCCAGAGGATGGCAGAGTCTATGCTCGTGGGATTATGGC 121
 QY 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
 DB 122 TATCAGATTACAGCTATGGCAACCCGAAATTTTAGTGGACAATAAACACATGGAAATAT 181
 QY 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIle 569
 DB 182 ACCATATTCTCGAGTAGGGCAGGCATGGCTCCACAGAGGAAAAGCCTATGAGATC 241
 QY 570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSer 589
 DB 242 ATGAGGAGCTCGATGTCAGCTATGTGCTGCTCATTTTGGAGGCTCTACTGGGTATTC 301
 QY 590 SerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGly 609
 DB 302 TCTGATGATATCAACAGATTCTTTGGATGTCGAGTGGGAGGACACAGATACAGGC 361
 QY 610 LysHisIleLysGluAsnAspTyrTyrProThrGlyGluPheArgValAspArgGlu 629
 DB 362 AAACATATCAAGGAGATGACTATTATCTCAACTGGGAGTTCGCTGTGGACCGTGAA 421

RESULT 14
 US-10-146-502-1456
 ; Sequence 1456, Application US/10146502
 ; Publication No. US20030069180A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Wang, Aijun
 ; APPLICANT: Stolk, John A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 ; FILE REFERENCE: 210121.527C2
 ; CURRENT APPLICATION NUMBER: US/10/146,502
 ; CURRENT FILING DATE: 2002-05-14
 ; NUMBER OF SEQ ID NOS: 2241
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1456
 ; LENGTH: 452
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 422
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-146-502-1456

Alignment Scores:
 Pred. No.: 1,028-137 Length: 452
 Score: 140.00 Matches: 140
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.86% Indels: 0
 DB: 14 Gaps: 0

US-10-028-384-12 (1-705) x US-10-146-502-1456 (1-452)

QY 490 LeuSerAlaArgGlyGlyAspGlySerArgGlyIlePheAspAspPheArgGluAlaTyr 509
 DB 2 CTATCTCCCGTGGTGGGATGGCAGTAGGATCATATTTGATGACTTCCGAGAAGCATAT 61
 QY 510 TyrTrpLeuArgHisAsnThrProGluAspAlaValMetSerTrpAspTyrGly 529
 DB 62 TATTGGCTTCGTATATATCTCCAGAGGATGGCAGAGTCTATGCTCGTGGGATTATGGC 121
 QY 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
 DB 122 TATCAGATTACAGCTATGGCAACCCGAAATTTTAGTGGACAATAAACACATGGAAATAT 181
 QY 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIle 569
 DB 182 ACCATATTCTCGAGTAGGGCAGGCATGGCTCCACAGAGGAAAAGCCTATGAGATC 241
 QY 570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSer 589
 DB 242 ATGAGGAGCTCGATGTCAGCTATGTGCTGCTCATTTTGGAGGCTCTACTGGGTATTC 301
 QY 590 SerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGly 609
 DB 302 TCTGATGATATCAACAGATTCTTTGGATGTCGAGTGGGAGGACACAGATACAGGC 361
 QY 610 LysHisIleLysGluAsnAspTyrTyrProThrGlyGluPheArgValAspArgGlu 629
 DB 362 AAACATATCAAGGAGATGACTATTATCTCAACTGGGAGTTCGCTGTGGACCGTGAA 421

RESULT 15
 US-09-871-161-213
 ; Sequence 213, Application US/09871161
 ; Publication No. US20030097866A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ENDEGE, WILSON O., ET AL.
 ; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 ; TITLE OF INVENTION: PRODUCTS: II
 ; FILE REFERENCE: CCDNA-260XX
 ; CURRENT APPLICATION NUMBER: US/09/871,161
 ; CURRENT FILING DATE: 2001-05-31
 ; PRIOR APPLICATION NUMBER: 09/328,111
 ; PRIOR FILING DATE: 1999-06-08
 ; PRIOR APPLICATION NUMBER: 60/117,393
 ; PRIOR FILING DATE: 1999-01-27
 ; PRIOR APPLICATION NUMBER: 60/098,639
 ; PRIOR FILING DATE: 1998-08-31
 ; NUMBER OF SEQ ID NOS: 544
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 213
 ; LENGTH: 487
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-09-871-161-213

Alignment Scores:
 Pred. No.: 5,168-125 Length: 487
 Score: 128.00 Matches: 128
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 18.16% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-12 (1-705) x US-09-871-161-213 (1-487)

```
QY 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeu 290
Db 88 TTTGGGTCCTTTGGTCTCTGCAGATCCATGCCCTTTGTGGATTACCTGGCAGCAAGTTG 147
QY 291 AsnProGlnPheGlnValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
Db 148 AATCCACACAAATTGAGTTCTTTTCGGAGCGTCATCTCTCGGAGGCTTTGTCCCT 207
QY 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPhe 330
Db 208 CTCACCGTGGAGCTCTCTCATGCTGACAGGAAAAATATCTCCCTGGACGGGCGTTTC 267
QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSer 350
Db 268 TACTCACCTGCTGGATCCCTCTTATGCTAAGAACCAACATCCCCCATCATGTCTCTGTGTCT 327
QY 351 GluHisGlnProThrThrTyrSerSerTyrTyrPheAspLeuGlnLeuValPheMet 370
Db 328 GAGCATCAGCCCAACACCTGGTCTCTCATCTATTTTTCACCTGCAGCTCCTCTCTTCATG 387
QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIle 390
Db 388 TTTCCAGTTGGCCCTCTATTACTGCTTTAGCAACCTGTCTGATGCCCGGATTTTATCATC 447
QY 391 MetTyrGlyValThrSerMetTyr 398
Db 448 ATGTATGGTGTGACCAGCATGTAC 471
```

Search completed: December 15, 2004, 15:28:24
Job time : 930.605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:58 ; Search time 5457.4 Seconds
(without alignments)
4707.376 Million cell updates/sec

Title: US-10-028-384-12
Perfect score: 705
Sequence: 1 MTKGFLRLSVKQDTLLKL.....HMLVRIYKYKDLNRLSLST 705

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32022875 seqs, 18219865908 residues

Word size: 1
Total number of hits satisfying chosen parameters: 65644297

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-O=/cgn2.1/USP20.spool/US10028384/runat.14122004.131646.6807/app_query.fasta.1.3740
-DB=EST -OFMT=FASTAP -SUFFIX=Oligo.rst -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@CGN.1.13960@runat.14122004.131646.6807 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -FPGAPOP=6
-FPGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hcc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|----------------------|
| 1 | 226 | 32.1 | 1345 | 3 | CR606556 | CR606556 full-length |
| 2 | 218 | 30.9 | 1145 | 5 | BX446691 | BX446691 BX446691 |
| 3 | 210 | 29.8 | 845 | 4 | BG678139 | BG678139 602625911 |
| 4 | 198 | 28.1 | 2118 | 9 | AV418284 | AV418284 Homo sapi |
| 5 | 183 | 28.0 | 770 | 4 | B1859687 | B1859687 603386275 |
| 6 | 179 | 25.4 | 686 | 1 | AU123167 | AU123167 AU123167 |
| 7 | 178 | 25.2 | 541 | 7 | CN265339 | CN265339 1700005999 |
| 8 | 178 | 25.2 | 692 | 1 | AI905439 | AI905439 RC-BT091 |
| 9 | 164 | 23.3 | 994 | 6 | CA487523 | CA487523 AGENCOURT |

| | | | | | | |
|----|-----|------|------|---|----------|--------------------|
| 10 | 161 | 22.8 | 484 | 5 | BX284031 | BX284031 |
| 11 | 160 | 22.7 | 733 | 6 | CR38051 | CR38051 4063016 B |
| 12 | 157 | 22.3 | 914 | 6 | CA454755 | CA454755 AGENCOURT |
| 13 | 155 | 22.0 | 615 | 1 | AI437149 | AI437149 fb38c11.Y |
| 14 | 155 | 22.0 | 627 | 2 | BE251342 | BE251342 601107886 |
| 15 | 155 | 22.0 | 743 | 5 | BU471816 | BU471816 603760287 |
| 16 | 155 | 22.0 | 760 | 5 | BU248580 | BU248580 603779426 |
| 17 | 155 | 22.0 | 776 | 7 | CK027267 | CK027267 AGENCOURT |
| 18 | 155 | 22.0 | 816 | 5 | BU200678 | BU200678 603101974 |
| 19 | 155 | 22.0 | 824 | 5 | CK028993 | CK028993 AGENCOURT |
| 20 | 155 | 22.0 | 987 | 5 | BU356020 | BU356020 603474684 |
| 21 | 155 | 22.0 | 1124 | 7 | CK030365 | CK030365 AGENCOURT |
| 22 | 154 | 21.8 | 927 | 5 | BU543738 | BU543738 AGENCOURT |
| 23 | 153 | 21.7 | 556 | 6 | CB363245 | CB363245 2F001-P00 |
| 24 | 152 | 21.6 | 639 | 6 | CA531158 | CA531158 60315606 |
| 25 | 152 | 21.6 | 694 | 4 | B1654677 | B1654677 603285595 |
| 26 | 152 | 21.6 | 935 | 6 | CB208848 | CB208848 AGENCOURT |
| 27 | 152 | 21.6 | 2103 | 9 | AY418286 | AY418286 Mus muscu |
| 28 | 152 | 21.6 | 2656 | 3 | BC028897 | BC028897 Mus muscu |
| 29 | 152 | 21.6 | 2730 | 3 | AK081547 | AK081547 Mus muscu |
| 30 | 152 | 21.6 | 3802 | 3 | AK030363 | AK030363 Mus muscu |
| 31 | 151 | 21.4 | 643 | 5 | BU290292 | BU290292 604164809 |
| 32 | 148 | 21.0 | 481 | 7 | CF931151 | CF931151 CF-06-R- |
| 33 | 147 | 20.9 | 576 | 4 | BMS11440 | BMS11440 J148C04.Y |
| 34 | 147 | 20.9 | 814 | 4 | B1684137 | B1684137 603306543 |
| 35 | 146 | 20.7 | 620 | 6 | CB426082 | CB426082 601223 MA |
| 36 | 145 | 20.6 | 465 | 1 | AJ692414 | AJ692414 AJ692414 |
| 37 | 145 | 20.6 | 626 | 5 | BQ526963 | BQ526963 NISC no19 |
| 38 | 145 | 20.6 | 641 | 1 | AL865732 | AL865732 AL865732 |
| 39 | 145 | 20.6 | 849 | 7 | CR409802 | CR409802 CR409802 |
| 40 | 145 | 20.6 | 860 | 7 | CR416594 | CR416594 CR416594 |
| 41 | 145 | 20.6 | 864 | 7 | CO788779 | CO788779 spleen ES |
| 42 | 145 | 20.6 | 941 | 7 | CR585625 | CR585625 CR585625 |
| 43 | 144 | 20.4 | 600 | 6 | CA527576 | CA527576 8043-93 M |
| 44 | 144 | 20.4 | 650 | 7 | CN265342 | CN265342 170005327 |
| 45 | 144 | 20.4 | 688 | 2 | BF348828 | BF348828 QV0-DT002 |

ALIGNMENTS

CR606556 1345 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CLOB0092D10 of Neuroblastoma of Homo sapiens (human).

ACCESSION CR606556
VERSION - GI:50487363
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1345)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 1345)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail: segref@genoscope.cns.fr - Web: www.genoscope.cns.fr)

REFERENCE 2 (bases 1 to 1345)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail: segref@genoscope.cns.fr - Web: www.genoscope.cns.fr)

COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
LOCATION/Qualifiers 1..1345
/organism="Homo sapiens"

FEATURES
source


```

Db      489 TTATACTCCAACTGGGAGTTCGGTGGACCGTGAAGGTTCTCCAGTCTGCTCAACTG 548
Qy      637 sLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaValArgPr 657
Db      549 CCTATGTACAAAGATGTTACTATCGCTTTGACAGGTTTACACAGAGCCAGCGTCC 608
Qy      657 oProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVa 677
Db      609 TCCAGGCTTTACCGCTGTCGGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGAICT 668
Qy      677 lLeuGluGlu 680
Db      669 CTGGAGGAA 678

RESULT 4
AY418284
LOCUS      2118 bp DNA linear GSS 17-DEC-2003
DEFINITION Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY418284
VERSION AY418284.1 GI:39774244
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Ranenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 2118)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Ranenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source location/Qualifiers
1..2118
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>2118
/gene="ITM1"
/locus_tag="HCM650c"

gene

ORIGIN
Alignment Scores:
Pred. No.: 5,08e-202 Length: 2118
Score: 198.00 Matches: 592
Percent Similarity: 96.73% Conservative: 0
Best Local Similarity: 96.73% Mismatches: 10
Query Match: 28.09% Indels: 20
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AY418284 (1-2118)
Qy      51 TyrPheAsnTyrArgThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsn 70
Db      151 TACTTTTAATATCGGACTACCAAGTTCCTGGCTGAGAGGGGTTTATAAATTCCATAAC 210
Qy      71 TrpPheAspArgAlaIatTyrProLeuGlyArgIleGlyGlyThrIleTyrPro 90
Db      211 TGGTTTATGACCGAGCCTGGTACCCCTTGGGACGAATCATTTGGAGGAACAATTTACCCA 270

```

```

Qy      91 GlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIle 110
Db      271 GGTTTAATGATCACCTCTGCTGCAATCTACCATGTACTCCATTTTCCACATCACCACATC 330
Qy      111 AspIleArgAsnValCysValPheLeuAlaProLeuPheSerSerPheThrSer-IleVa 130
Db      331 GACATTGGGAATGTGTGTGTCTCTGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 389
Qy      130 lThrTyrLeu-LeuThrLysGluLeuLysAspAlaGlyValaGlyLeuLeuAlaAlaAla 150
Db      390 CAGGTACCA-CCTTACCAAGAGCTCAAGATGCAGGGGCTGGGCTTCTTGTCTGCCA 448
Qy      150 etIleAlaValValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGlu 170
Db      449 TGATTGCTGTAGTCTCTGGATATATCTCCGATCTGTGGCTGGCTCTGTATGATAAAG 508
Qy      170 lYleAlaIlePheCysMetLeuLeuThrTyrTyrMetTyrIleValaValaValaVala 190
Db      509 GGATTGCCATCTTTGGCATGCTACTACCTACTACATGTGATCAAGCGAGTAAGACTG 568
Qy      190 lYserIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTrp 210
Db      569 GTTCCATCTGTGGGAGCTAAGTGTGCCCTTGTCTATTTCTACATGGTCTCGTCATGG 628
Qy      210 lYgLYTyrValPheLeuIleAsnLeuIleProLeuHisValLeuValLeuMetLeuThr 230
Db      629 GAGGTTATGTGTCTCTGATCACTTAATTCCTCTCCAGTCTCTCTCTCTCTCTCTCTCA 688
Qy      230 lYargPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIle 250
Db      689 GCGGTTTCTCACCGGATCTATGTGGCTACTGTACTGTACTGTACTGTACTGTACTGT 748
Qy      250 euSerArg-GlnIleSerPheValGlyPheGlnProValLeuSerSerGluHisMetAla 269
Db      749 TTTCTAT-CCAGATCTCTTTGGGTTTCCAGCTGTCTCTCTCTCTCTCTCTCTCTCA 807
Qy      270 Gly-PheGlyValPheGlyLeuCysGlnIleHisAlaPheValaValaValaVala 289
Db      808 GC-CTTTGGGTTCTTTGGTCTCTGCCAGATCCATCCCTTTGTGGATTACCTCGCGAG 866
Qy      289 sLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheVa 309
Db      867 GTTGAATCCACAACTTTGAAGTCTTTCCGGAGGCTCATCTCTCTGTAGGCTTTGT 926
Qy      309 lLeuLeuThrValGlyAlaLeuLeuMetLeuThrGly----LysIleSerProThrThr 328
Db      927 CCTTCTCACCGTGGGAGCTCTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 982
Qy      328 lYargPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAla 348
Db      983 GGGGTTTCTACTCGTGTGGTGGATCCCTCTTATGTAGTAAAGAACATCCCATCTGCTT 1042
Qy      348 erValSerGluHisGlnProThrThrTrpSerSerTyrTyrPheAspLeuGlnLeuVal 368
Db      1043 CTGTGTCTGAGCATCAGCCCAACCTGGTCTCTCATCTATTTTGACCTGACGCTCTCT 1102
Qy      368 alPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIle 388
Db      1103 TCTTCATGTTTCCAGTGGCTCTCTTATCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1162
Qy      388 heIleIleMetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMet 408
Db      1163 TTATCATCATGTATGTGTGACCAAGCTGTATCTTTTTCAGTGTATGTGTGTGTGTGTGT 1222
Qy      408 euValLeuAlaProValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSer 427
Db      1223 TAGTGTGGCACTGTATGTG-CATCTCTCTGGCATTTGGAGTCTCTCCAGGTCTGTCTC 1281
Qy      428 ThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSerLysGlnGlnAsp 447
Db      1282 ACATACATGAAGAATCTGGACATAAGTCTGCACAGCAAGAACAGAACAGACAGGAT 1341

```

QY 448 SerThrTyrProIleLysIle-GluValAlaSerGlyMetIleLeuValMetIlePhe 467
 Db 1342 TCCACCTACCTATTAAAG-TGAAGTGGCAAGTGGGATCATCTGGTCATGCTTCTT 1400
 QY 467 eLeuIleThrTyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSe 487
 Db 1401 TCTCATCACCTACACCTTTCATCAACCTGGGTGACCACTGAGGCTTCTTCTCGTC 1460
 QY 487 rIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArg 507
 Db 1461 CATGTGATCTCTCGCCGTGGTGGGATGGCAGTGGATCATATTTGATGACTTCGAGA 1520
 QY 507 uAlaTyrTyrTyrIleuArgHisAsnThrProGluAspAlaLysValMetSerTyrPhe 527
 Db 1521 AGCATATTATTGGCTTCGTCTATACTATCTCCAGAGGATGCGAAGTCTATGCTCGTGG 1580
 QY 527 pTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnThrTyr 547
 Db 1581 TTATGGCTATCAGATTACAGTATGGCAACCCGAAACAAATTTTAGTGGACAATAACACATG 1640
 QY 547 pAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGlyLysAlaTyr 567
 Db 1641 GAATAATACCCATATTTCTCGATGAGGAGGCAATGGCGTCCACAGAGGAAAGCCCTA 1700
 QY 567 rGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrG 587
 Db 1701 TGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTTGGAGGCTCTACTGG 1760
 QY 587 yTyrSerSerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAs 607
 Db 1761 GTATTCCTCTCATGATATCAACAAGTTCTTTGGATGCTCGGATTTGGAGGAGCACAGA 1820
 QY 607 pThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAs 627
 Db 1821 TACAGGCAACATATCAAGGAGATGACTATTATCTCAACCTGGGAGTTCGCTGTGGA 1880
 QY 627 pArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPh 647
 Db 1881 CCGTAGAGTTCCTCAGTGTCTCACTGCTCACTGCTCATGCTACAGATGTGTACTATCGCTT 1940
 QY 647 eGlyGlnValTyrThr 652
 Db 1941 TGGACAGGTTTACACA 1956

RESULT 5
 BI859687
 LOCUS 603388275F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5394834 5',
 DEFINITION mRNA sequence.
 ACCESSION BI859687
 VERSION BI859687.1 GI:16000434
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (Bases 1 to 770)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHAM12006 row: f column: 19
 High quality sequence stop: 620.
 Location/Qualifiers
 1..770

FEATURES
 source

RESULT 6
 AU123167
 LOCUS AU123167 NT2RM1 Homo sapiens cDNA clone NT2RM1000838 5', mRNA
 DEFINITION AU123167 NT2RM1 Homo sapiens cDNA clone NT2RM1000838 5', mRNA
 sequence.

ORIGIN
 Alignment Scores: 3.02e-186 Length: 770
 Pred. No.: 183.00 Matches: 209
 Score: 99.05% Conservative: 0
 Percent Similarity: 99.05% Mismatches: 0
 Best Local Similarity: 99.05% Indels: 2
 Query Match: 25.96% Gaps: 0
 DB: 4
 US-10-028-384-12 (1-705) x BI859687 (1-770)

QY 462 LeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrValThrSerGlu 481
 Db 2 CTGGTCATGCTTCTTCTTCATCACCCTACACCTTTTCATTCACCTGGGTGACCACTGAG 61
 QY 482 AlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIle 501
 Db 62 GCCTACTCTTCTCGTCCCTGCTTACTATCTGCGCGTGGTGGGATGGCAGTAGGATCAT 121
 QY 502 PheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLys 521
 Db 122 TTTGATCACTTCCGAGAGCATATATTGGCTTCGTTCATAAATACCTCCAGAGTAGCAAG 181
 QY 522 ValMetSerTyrThrPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
 Db 182 GTCATGTCCTGGTGGGATTTATGCTATCAGATTACAGCTATGGCAACCGAACAAATT 241
 QY 542 ValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
 Db 242 GTGGACAATAACACATGGAATAATATCCCATATTTCTCGAGTAGGGCAGCAATGGCTCC 301
 QY 562 ThrGluGlyLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIle 581
 Db 302 ACAGAGGAAAAGCCCTATGAGATCATGAGGAGCTCGATGTGCTGATGTGCTGTCATT 361
 QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIle-AsnLysPheLeuTyrMetValAr 601
 Db 362 TTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAAAACAAGTTTCTTTGGATGGTCCG 421
 QY 601 gIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProTh 621
 Db 422 GATTGGAGGAGCACAGATACAGGCAACATATCAGAGGAGATGACTATTATCTCCACAC 481
 QY 621 rGlyGluPheArgValAspArgGlySerProValLeuLeuAsnCysLeuMetTyrIly 641
 Db 482 TGGGAGTTCCGTGTGGACCGTGAAGGTCTCTCCAGTGTGCTCAACTGCTCATGTACAA 541
 QY 641 sMetCysTyrTyrArgPheGlyGlnValTyrGlu-AlaLysArgProGlyPheA 661
 Db 542 GATGTGTATTATCGCTTTGGACAGTTTACACAGAAAGCGGCTCTCCAGGCTTTG 601
 QY 661 spArgValArgAsnAlaGluIleGlyAsn 670
 Db 602 ACCGTGTCCGAAATGCTGAGATCGGAAT 630

US-10-028-384-12 (1-705) x BI859687 (1-770)
 Alignment Scores: 3.02e-186 Length: 770
 Pred. No.: 183.00 Matches: 209
 Score: 99.05% Conservative: 0
 Percent Similarity: 99.05% Mismatches: 0
 Best Local Similarity: 99.05% Indels: 2
 Query Match: 25.96% Gaps: 0
 DB: 4
 US-10-028-384-12 (1-705) x BI859687 (1-770)

```

ACCESSION      AUI231167
VERSION        AUI231167.1  GI:10947883
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE     1 (bases 1 to 686)
AUTHORS      Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y., Ishii,S., Saito,K., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.
TITLE        HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Kawai,Y., Ishii,S., Saito,K., Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.)
JOURNAL      Unpublished (2000)
COMMENT      Contact: Takao Isogai
              Genomics Laboratory
              Helix Research Institute
              1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              Tel: 81-438-52-3975
              Fax: 81-438-52-3986
              Email: genomics@hri.co.jp
              HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
              Research Institute; cDNA library construction; Department of
              Virology, Institute of Medical Science, University of Tokyo, and
              Helix Research Institute
FEATURES      Location/Qualifiers
              source
                1..686
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="NT2RM1000838"
                /cell_type="teratocarcinoma"
                /cell_line="NT2"
                /clone_lib="NT2RM1"
                /notes="Vector: pUC19FL3; mRNA from uninduced NT2 neuronal
                precursor cells"
ORIGIN
Alignment Scores:
Pred. No.:      5,65e-182      Length:      686
Score:          179.00      Matches:      179
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     25.39%      Indels:      0
DB:             1      Gaps:      0
US-10-028-384-12 (1-705) x AUI231167 (1-686)
QY 502 PheAspAspPheArgGluAlaTyrTyrTyrTyrLeuArgHisAsnThrProGluAspAlaLys 521
DB 3 TTGTGATGACTTCGAGAGACCAATATTATGCTTCGTATPACATCCAGAGATCGGAG 62
QY 522 ValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
DB 63 GTCATGTCTCCGTGGGATTATGCTATCAGATTACAGCTATGCGAAACCGAACAATTTA 122
QY 542 ValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
DB 123 GTGGACATTAACATGGAATATATACCATATTTCTCGAGTAGGACGAGCATGGCGTCC 182
QY 562 ThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleuValile 581
DB 183 ACAGAGAAAGACCTATGAGATCATGAGGAGAGCTCGATGTCAGCTATGCTGCTCAT 242
QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArg 601
DB 243 TTGGAGCCCTCACTGGGTATCTCTGATGATATACCAAGTTCTTGTGGATGGTCCGG 302
QY 602 IleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrProThr 621
DB 303 ATTGGAGGAGCAGATACAGCAACATATCAAGGAGAAATGACTATATTACTCCAACT 362
QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
363 GGGGAGTTCCGTGTGGACCGCTGAAGGTTCTCCAGTGTGCTCACTGCCTCATGTACAAG 422
642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
423 ATGTGTTACTATCGCTTTGGACAGGTTTACACAGAACCAAGGCTCTCCAGGCTTTGAC 482
662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
483 CGTGTCCCAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGTCTCTGGAGGAA 539
RESULT 7
CN265339
LOCUS      CN265339
DEFINITION 17000599935691 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION  CN265339
VERSION    CN265339.1  GI:47281753
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 541)
AUTHORS    Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J., Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R., Lebkowski,J and Stanton,L.W
TITLE      Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation
JOURNAL    Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT    Contact: Brandenberger R
              Regenerative Medicine
              Geron Corporation
              230 Constitution Drive, Menlo Park, CA 94025, USA
              Tel: 650 473 8658
              Fax: 650 473 7760
              Email: rbrandenberger@geron.com
              Insert Length: 541 Std Error: 0.00.
              Location/Qualifiers
                1..541
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /tissue_type="embryonic stem cell, retinoic acid and
                mitogen-treated hES cell line H7"
                /clone_lib="GRN_PRENEU"
                /notes="cligo dt primed, full-length enriched cDNA library
                from hES cell line H7 (p29) maintained in feeder-free
                conditions. Embryoid bodies were generated in the presence
                of all-trans retinoic acid and mitogens."
ORIGIN
Alignment Scores:
Pred. No.:      5,39e-181      Length:      541
Score:          178.00      Matches:      178
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     25.25%      Indels:      0
DB:             7      Gaps:      0
US-10-028-384-12 (1-705) x CN265339 (1-541)
QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHis 474
DB 6 GAAGTGGCAAGTGGGATGATGACTGCTCATGCTGCTTTCTTCTCATCACCTACACTTCAT 65
QY 475 SerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494
DB 66 TCACCTGGGTGACAGTAGGAGGCTACTCTCTCCGTCATTGATCTATCTGCCCCGTGT 125
QY 495 GlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHis 514
DB 126 GGGGATGGCAGTAGGATCATATTGATGACTTCGAGAGCATATTATTGGCTTCGTCAT 185
QY 515 AsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAla 534

```

```

Db 186 AATACTCCAGAGGATCGAAGTCTATGCTGGTGGATTATGCTATCAGATTACAGCT 245
Qy 535 MetAlaAsnArgThrIleValAspAsnSerThrTrpAsnSerThrHisIleSerArg 554
Db 246 ATGGCAACCCGACCAATTTTATGACACATACACATGGAATATATCCCATATTTCTCGA 305
Qy 555 ValGlyGlnAlaMetAlaSerThrGluGluLysAlaValTrpGluLeuMetArgGluLeuAsp 574
Db 306 GTAGGCGCAGGAATGGCGTCACAGAGGAAAAGCCTATGAGATCATGAGGAGGCTCGAT 365
Qy 575 ValSerTrpValLeuValIlePheGlyValLeuThrGlyTrpSerSerPheAspIleAsn 594
Db 366 GTACACTATGCTGGTCAATTTTGGAGGCTCTACTGGTATTCCTCTGATGATATCAAC 425
Qy 595 LysPheLeuTrpMetValArgIleGlySerThrAspThrGlyHisIleLeuLysGlu 614
Db 426 AAGTTTCTTTGGATGGTCCGATTTGGAGGAGGACAGATACAGGCAACATATCAAGGAG 485
Qy 615 AsnAspTyrTrpThrProThrGlyGluPheArgValAspArgGluGlySerPro 632
Db 486 AATGACTATTTACTCCCACTGGGAGTTCCGTGGACCGTGAAGGTTCTCCA 539

RESULT 8
AI905439
LOCUS RC-BT091-200199-026 BT091 Homo sapiens cDNA, linear EST 30-MAR-2000
DEFINITION RC-BT091-200199-026 BT091 Homo sapiens cDNA, mRNA sequence.
ACCESSION AI905439
VERSION AI905439.1 GI:6495826
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 692)
Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jorgensen, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/seq/gethtml.pl?t1=RC&t2=RC-BT091-026.html
&t3=200199&t4=1)
Seq primer: puc 18 forward.
Location/Qualifiers
1..692
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="female"
/dev_stage="Adult"
/clone_lib="BT091"
/note="Organ: breast; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No.
196,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low

```

```

stringency conditions."
ORIGIN
Alignment Scores:
Pred. No.: 6.95e-181 Length: 692
Score: 178.00 Matches: 178
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.25% Indels: 0
DB: 1 Gaps: 0
US-10-028-384-12 (1-705) x AI905439 (1-692)
Qy 476 ThrTrpValThrSerGluAlaValTrpSerSerProSerIleValLeuSerAlaArgGlyGly 495
Db 20 ACCGTGGGTGACCATGAGGCTACTCTCTCCGTCCATTGTACTATCTCCCGTGGTGG 79
Qy 496 AspGlySerArgIleIlePheAspAspPheArgGluAlaValTrpTrpLeuArgHisAsn 515
Db 80 GATCGCAGTAGGATCATATTTGATGACTCCGAGAGCATATTTATGGCTTCGTATAT 139
Qy 516 ThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTrpGlnIleThrAlaMet 535
Db 140 ACTCCAGAGGATCGAAGGTCATGCTCTGCTGGGATTTATGGCTATCAGATTACAGTATG 199
Qy 536 AlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgVal 555
Db 200 GCAACCCGACCAATTTTATGAGACATATACACATGGAATATATACCATATTTCTCGAGTA 259
Qy 556 GlyGlnAlaMetAlaSerThrGluGluLysAlaValTrpGluLeuMetArgGluLeuAspVal 575
Db 260 GGGCAGCAATGGCGTCCACAGAGGAAAAGCCTATGAGATCATGAGGAGGCTCGATGTC 319
Qy 576 SerTrpValLeuValIlePheGlyGlyLeuThrGlyTrpSerSerAspAspIleAsnLys 595
Db 320 AGCTATGCTGGTCAATTTTGGAGGCCCTACTGGGTATTCCTCTGATCATATCAACAG 379
Qy 596 PheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsn 615
Db 380 TTTCTTTGGATGGTTCGGGATTTGGAGGAGCAGACAGATACAGCAACATATCAAGGAGAT 439
Qy 616 AspTrpTrpThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeu 635
Db 440 GACTATTTATCTCCACTCGGAGTTCGGTGGACCGTGAAGGTTCTCCAGTGTGCTC 499
Qy 636 AsnCysLeuMetTyrLysMetCysTrpTrpArgPheGlyGlnValTrpThrGlu 653
Db 500 AACTGCCTCATGTACAGATGTGCTACTATCGCTTTTGGACAGGTTTACACAGAA 553

RESULT 9
AI905439
LOCUS RC-BT091-200199-026 BT091 Homo sapiens cDNA clone IMAGE:6719013 5',
DEFINITION mRNA sequence.
ACCESSION CA487523
VERSION CA487523.1 GI:24947110
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 994)
NHL-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbe@mail.nih.gov
Tissue Procurement: Kristi A. Egland, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

```


http://image.llnl.gov
 Plate: LAM14276 row: d column: 21
 High quality sequence stop: 623
 Location/Qualifiers

FEATURES

source

1..994
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="IMAGE:6719013"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 HTERT-HME1, LNCaP"
 /lab_host="EVDH10B"
 /lab_host="EVDH10B"
 /note="vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: Oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Sungkook Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:
 Pred. No.: 1.29e-165 Length: 994
 Score: 164.00 Matches: 164
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 23.26% Indels: 0
 DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x CA487523 (1-994)

QY 517 ProGluAspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAla 536
 DB 1 CCAGAGATCGGAGGTCATGCTCTGGTGGATTATGGCTATCAGATTACAGCTATGCA 60
 QY 537 AsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGly 556
 DB 61 AACCGAACATATTTAGTGGCAATAACACATGGATTAATACCCATATTTCTGAGTAGGG 120
 QY 557 GlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgLysLeuAspValSer 576
 DB 121 CAGGCATGGCTCCACAGAGGAAAGGCTATGATCATGAGGAGCTCGATGTCAGC 180
 QY 577 TyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPhe 596
 DB 181 TATGTCTGTGTCATTTTGGAGGCTCTCTGGTATTCCTCTGATGATATCAACAAGTTT 240
 QY 597 LeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAsp 616
 DB 241 CTTTGGATGTCGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGAC 300
 QY 617 TyrTrpThrProThrGlyGlyPheArgValAspArgGlyGlySerProValLeuLeuAsn 636
 DB 301 TATTATCTCAACTGGGGAGTTCCGTGTGGACCGTGAAGGTTCTCCAGTGTCTGCTAAC 360
 QY 637 CysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArg 656
 DB 361 TGCCTCANGTACAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCCGAGGCT 420
 QY 657 ProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheCylLeuAsp 676
 DB 421 CCTCCAGGCTTTCACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGA 480
 QY 677 ValLeuGluGlu 680
 DB 481 GTCCTGGAGGA 492

RESULT 10

BX284031

LOCUS

DEFINITION BX284031 NCI_CGAP_skn4 Homo sapiens cDNA clone IMAGE998E0410606 ;

484 bp

mRNA

linear

EST 05-MAR-2003

IMAGE:4750827, mRNA sequence.

ACCESSION BX284031

VERSION BX284031.1 GI:28848485

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 484)

AUTHORS

Ebert, L., Heil, O., Hennig, S., Neubert, P., Partsch, E., Peters, M.,

TITLE

Human Unigeneset - RZPD3

JOURNAL

Unpublished (2003)

COMMENT

Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD; IMAGE998E0410606.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Human Unigeneset - RZPD3 (RZPDLIB No.972)

http://www.rzpd.de/CloneCards/cgi-

bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 101

Fax: +49 30 32639 111

www.rzpd.de

This clone is available royalty-free from RZPD;

Contact RZPD (clone@rzpd.de) for further information. Seq primer:

SP6, Primer sequence: ATTAGTGACACTATAG.

FEATURES

source

1..484

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE998E0410606 ; IMAGE:4750827"

/tissue_type="squamous cell carcinoma"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NCI_CGAP_skn4"

/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: oligo dr.

Average insert size 1.5kb. Library constructed by Life

Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Alignment Scores:

Pred. No.: 1.11e-162 Length: 484
 Score: 161.00 Matches: 161
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 22.84% Indels: 0
 DB: 5 Gaps: 0

US-10-028-384-12 (1-705) x BX284031 (1-484)

QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPheLeuIleThrTyrPheHis 474
 DB 1 GAAGTGGCAAGTGGGATGATGATCTGGTCATGGCTTCTTCTCATCCTACCTACCTTCAT 60
 QY 475 SerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494
 DB 61 TCACCTGGGTGGACCAAGGAGGCTCTCTCTCCGTCCTTGTACTATCTGCCGTGGT 120
 QY 495 GlyAspGlySerArgIlePheAspAspPheArgGluAlaTyrTrpLeuArgHis 514
 DB 121 GGGGATGGCAGTGGATCATATTTGATGACTTCCGAGACATATTTATGGCTTCGCTCAT 180
 QY 515 AsnThrProGluAspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAla 534
 DB 181 AATATCTCCAGAGGATGCCAAGGTCATGTCCTGGTGGGATTTATGGCTATCAGATTACAGCT 240
 QY 535 MetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArg 554
 DB 241 ATGGCAACCCGACAAATTTTATGTGGACAATAACATGGAATTAATACCCATATTTCTCGA 300

```

QY 555 ValGlyGlnAlaMetAlaSerThrCluGluLysAlaTyrGluLleMetArgGluLeuAsp 574
Db 301 GTAGGCGCAGGAGTGGCTGCACAGAGGAAAGCCATGAGATCATGAGGGAGCTCGAT 360
QY 575 ValSerTyrValLeuValLysPheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsn 594
Db 361 GTACAGTATGTGCTGGTCAATTTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAAC 420
QY 595 LysPheLeuTyrMetValArgGlyGlySerThrAspThrGlyLysHisIleLysGlu 614
Db 421 AAGTTTCTTTGGATGGTCCGATTTGGAGGAGGACAGATACAGCAAAATATCAAGGAG 480
QY 615 Asn 615
Db 481 AAT 483

RESULT 11
LOCUS CK838051
DEFINITION 733 bp mRNA linear EST 04-MAR-2004
sequence.
ACCESSION CK838051
VERSION CK838051.1 GI:45071891
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus

REFERENCE
AUTHORS Baumann,R.G., Baldwin,R.L., Sonstegard,T.S., Van Tassell,C.P. and
Matukumalli,L.K.
TITLE Construction and Analysis of a cDNA Library Generated From
Intestinal Muscle and Epithelial Tissues of Holstein Cattle
JOURNAL Unpublished (2004)
COMMENT Contact: Richard G. Baumann
Bovine Functional Genomics Lab
ANRI
BLDG 162: BARC-EAST, Beltsville, MD 20705, USA
Tel: 3015048604
Fax: 3015048744
Email: rbaumann@nri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim.alt -trim.fasta. Vector identified
by cross match using options -minmatch 12 -minscore 12
Plate: 10 row: J column: 04
Seq primer: CCTATTAGGTGACACTATAGAAC
High quality sequence stop: 733.

FEATURES
Source
1..733
/organism="Bos taurus"
/mol_type="mRNA"
/strain="Holstein"
/db_xref="taxon:9913"
/clone="8BOV 10J04"
/sex="female"
/tissue_type="Epithelial, Muscle"
/dev_stage="Lactating, Neonatal"
/lab_host="DH10B TONa"
/clone_lib="BARC 8BOV"
/Note="Organ: Intestine; Vector: pCMVSPORT6.1; Site: 1:
NotI; Site 2: EcoRI; Normalized cow cDNA intestinal
library in pCMVSPORT6.1, constructed from equimolar mRNA
pools derived from 5 sources: 4 lactating intestinal, 1
neonatal intestinal 4/5 Lactating, proximal Duodenum,
Jejunum, Distal ileum, Colon, 1/5 Neonatal, Proximal
Duodenum, Jejunum, Distal ileum"

ORIGIN
Alignment Scores:
Pred. No.: 733
Score: 160.00
Matches: 160

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 22.70% Indels: 0
DB: Gaps: 0

US-10-028-384-12 (1-705) x CK838051 (1-733)
QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuThrTyrThrPheHis 474
Db 203 GAAGTGGCAAGTGGCATGATGATGCTTCTTCTCATTTACCTACACCTTCCAC 262
QY 475 SerThrTyrValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494
Db 263 TCGACCTGGGTGACCAAGTGGGCTACTCTTCTCCCTCCATTGTGCTGCTGTGT 322
QY 495 GlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHis 514
Db 323 GGAGATGGCAGTAGGATCATTTTGTGATTTTCGAGAAGCGTACTATTGGCTTCGTCA 382
QY 515 AsnThrProGluAspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAla 534
Db 383 AATACTCCAGAGGATGCGAAGGTCAATGCTGGGATTTATGGCTACCAAGATTACAGCT 442
QY 535 MetAlaAsnArgThrIleLeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArg 554
Db 443 ATGGCGAATCGGACGATTTTAGTGGATAATAACACCTGGATAATACCATATATCTCGA 502
QY 555 ValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluLleMetArgGluLeuAsp 574
Db 503 GTAGGCGCAGCAATGGCATCCACAGAGAAAGAAAGCTATGAGATCATGAGGGAGCTTGAT 562
QY 575 ValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsn 594
Db 563 GTACAGTATGCTGGTGCATTTTGGAGGCTCACTGGGTATTTCTTCAGATGACATCAAC 622
QY 595 LysPheLeuTyrMetValArgGlyGlySerThrAspThrGlyLysHisIleLysGlu 614
Db 623 AAATTTCTGTGGATGGTCCGGATTGGAGGAGGACAGATACAGAGGAAACACATCAAGGAG 682

RESULT 12
LOCUS CA454755
DEFINITION 914 bp mRNA linear EST 12-NOV-2002
AGENCOURT_10763160 MAPCL Homo sapiens cDNA clone IMAGE:6721291 5',
mRNA sequence.
ACCESSION CA454755
VERSION CA454755.1 GI:24904799
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1. (bases 1 to 914)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Kristi A. Egland, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM14282 row: c column: 19
High quality sequence stop: 687.
Location/Qualifiers
1..914
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6721291"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,"

FEATURES
source

```

HTERT-HMEL, LNCaP"
/lab_host="EMD110B"
/clone_lib="WAPCL"
/notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kirsti A. Eglund, James J. Vincent, Robert Strausberg,
Bungkok Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 4 31e-158 Length: 914
Score: 157.00 Matches: 204
Percent Similarity: 99.03% Conservative: 0
Best Local Similarity: 99.03% Mismatches: 1
Query Match: 22.27% Indels: 2
DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x CA454755 (1-914)

QY 476 ThrTrpValThrSerGluAlaTyrSerProSerIleValLeuSerAlaArgGlyCly 495
Db 36 ACCTGGGTGACCAAGTGAAGGCTACTCTTCCTGCTCCATGTACTATCTCCCGGTGGG 95
QY 496 AspGlySerArgIleLeuPheAspPheArgGluAlaTyrTrpLeuArgHisAsn 515
Db 96 GATGCAAGTAGGATCATATTGATGACTTCCGAGAGCATATATTGCTTCGTCTAAT 155
QY 516 ThrProGluAspAlaValMet-SerTrpTrpAspTyrGlyTyrGlnIleThrAlaMe 535
Db 156 ACTCCAGAGATCGAAGGCTT-TGTCTGGTGGGATTATGCTATCATGATTACAGCTAT 214
QY 535 tAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgVa 555
Db 215 GGCAACCGAACAAATTTAGTGACCAATAACACATGGATATATACCATATTTCTCGAGT 274
QY 555 lGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspVa 575
Db 275 AGGCAGGCAATGGCGTCCACAGAGGAAAGCCTATGAGATCATGAGGAGGCTCGATGT 334
QY 575 lSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLy 595
Db 335 CAGCTATGTCTGCTCATTTTGGAGGCTCTACTGGTATTCCTCTGATGATATCAACA 394
QY 595 sPheLeuTrpMetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAs 615
Db 395 GTTTCCTTGGATGGTCCGATTGAGGAGGACACAGATACAGCAACATATACAGGAGAA 454
QY 615 nAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLe 635
Db 455 TGACTATTATCTCACTGGGAGTTCGGTGGACCGTGAAGGTTCTCCAGTGTCT 514
QY 635 uAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLy 655
Db 515 CAACTGGCTCATGTACAAAGATGTGTACTATGCTTTGACAGGTTTACACAGAGCCAA 574
QY 655 sATGProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLe 675
Db 575 GCGTCTCCAGGCTTTGACGCTGCGAAATGCTGAGATTGGATATAGACTTTGAGCT 634
QY 675 uAspValLeuGluGlu 680
Db 635 TGATGTCTCGAGGAA 650

RESULT 13

AI437149 615 bp mRNA linear EST 07-JUN-2001
LOCUS
DEFINITION
f38c11.v1 Zebrafish Washu MPIMG EST Danio rerio cDNA clone
IMAGE:37141645 similar to SW:STT3_MOUSE P46978 OLIGOSACCHARYL
TRANSFERASE STT3 SUBUNIT HOMOLOG 1; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI437149
AI437149.1 GI:4288291

EST.
Danio rerio (zebrafish)

Danio rerio
Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE
AUTHORS
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Thaising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
Washu Zebrafish EST Project 1998

Unpublished (1998)

CONTACT: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourceCenterPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)

Possible reversed clone: similarity on wrong strand

Seq primer: T3 ET from Amersham

High quality sequence stop: 497

POLYA=No.

FEATURES

source

1. 615
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:3714164"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/clone_lib="Zebrafish Washu MPIMG EST"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
15'pGACTAGTCTAGATCGGAGCGCCGCTTTTCTTTTCTTTT3';
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridisation fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

ORIGIN

Alignment Scores:

Pred. No.: 4.24e-156 Length: 615
Score: 155.00 Matches: 155
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.99% Indels: 0
DB: 1 Gaps: 0

US-10-028-384-12 (1-705) x A1437149 (1-615)

QY 455 GluValAlaSerGlyMetIleLeuValMetAlaPheLeuIleThrTyrThrPheHis 474

DB 49 GAAGTGGCAAGTGGGATGATCTAGTATGGGTTCTCTCATCATATACACTTCCAC 108

QY 475 SerThrTyrValThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGly 494

DB 109 TCCACCTGGGTGACAGCGGCGGTATTCATCCCTCCATTCCTGCTTCTGCTGGCGGT 168

QY 495 GlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHis 514

DB 169 GCGAGCGGCAGCGCATCATCTTTGATGACCTTTCAGAGAGGCTTATCTGCTGCTGACAC 228

QY 515 AsnThrProGluAspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAla 534

DB 229 AACACACAGAGAGTGTAAAGTATGCTGCTGGTGGGATTATGGTATCATAGATAACAGCG 288

QY 535 MetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArg 554

DB 289 ATGCCCAATCGAACGATCTGGTTGACATTAACACTTGGGAATACACTCATCTCCAGA 348

QY 555 ValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAsp 574

DB 349 GTGGTTCAGGCAATGGCTCCACAGAGAGAGAAAGCCTATGAGATTATGGAGAGCTGGAT 408

QY 575 ValSerTyrValLeuValIlePheGlyLeuThrGlyTyrSerSerAspAspIleAsn 594

DB 409 GTGAGCTACGCTCCGTGGTATCTTTGGCGGATTACCGGATATTCATCTGATGATATAAT 468

QY 595 LysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGly 609

DB 469 AAGTTCCTGCTGGTGGTGGTATCGGTGGGAGACACAGAGGG 513

RESULT 14

LOCUS BE251342

DEFINITION 60110786F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3344036 5', mRNA sequence.

ACCESSION BE251342

VERSION BE251342.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 627)

TITLE NIH-MGC http://mgi.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLC129 row: 1 column: 21

High quality sequence stop: 608.

Location/Qualifiers

1. 627

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3344036"

/tissue_type="retinoblastoma"

/lab_host="PH108 (phage-resistant)"

/clone_lib="NIH_MGC_16"

/note="Organ: eye; Vector: pOT7; Site:1; XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally-cloned into EcoRI/XhoI sites using the following 5'

FEATURES

source

adaptor: GGCACGAG (G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 4,32e-156 Length: 627

Score: 155.00 Matches: 188

Percent Similarity: 98.95% Conservative: 0

Best Local Similarity: 98.95% Mismatches: 1

Query Match: 21.99% Indels: 2

DB: 2 Gaps: 0

US-10-028-384-12 (1-705) x BE251342 (1-627)

QY 421 GlyValSerGlnValLeuSerThrTyrMetIleValLeuSerArgProAspLys 440

DB 1 GGAGTCTCCAGGTCGTGTCACATACATGAAGATCTGGACATAAGTCGTCACAGACAG 60

QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaSerGlyMet 460

DB 61 AAGACCAAGACACACAGATTCACCTACCTATTAAGAA-TGAAGTGGCAAGTGGGAT 119

QY 460 IleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTyrValThrSe 480

DB 120 GATACCTGGTCTGCTGCTTCTTCTCATCACCTACACCTTTCATCACTGGGTGACAC 179

QY 480 GCluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgI 500

DB 180 TGAGCCCTACTCTCTCCGTCATGTACTATCTGCCGCTGGTGGGATGTCAGTAGGAT 239

QY 500 eIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAl 520

DB 240 CATATTTGATGACATTCGAGAGACATATATTGCTTCGTCATATATCTCCAGAGGATGC 299

QY 520 alyValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrI 540

DB 300 GAAGTCTATGCTCTGGTGGGATTAATGGCTATCAGATACAGCTATGCGCAACCAAT 359

QY 540 eLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAl 560

DB 360 TTTAGTGACATTAACATGGAATATATACCATATTTCTCGATAGGAGGCAATGGC 419

QY 560 aSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVa 580

DB 420 GTCCACAGAGAGAAAGCCATGAGATCATGAGGAGCTCGATGTCAGCTATGTGCTGGT 479

QY 580 lilePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVa 600

DB 480 CATTTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAACAAGTTCTTTGGATGGT 539

QY 600 lArgIleGlyGlySerThrAspThrGly 609

DB 540 CCGGATTGGAGGAGCAGACATACAGGC 567

RESULT 15

LOCUS BU471816

DEFINITION 603760287F1 CSQRN21 Gallus gallus cDNA clone CHEST677n22 5', mRNA sequence.

ACCESSION BU471816

VERSION BU471816.1

KEYWORDS EST.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

AUTHORS 1 (bases 1 to 743)

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 0161208930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 743

FEATURES

source
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="layer"
/db_xref="taxon:9031"
/clone="CHEST677n22"
/sex="female"
/dev_stage="adult"
/lab_host="DH103"
/clone_lib="CSEORBN21"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN

Alignment Scores:
Pred. No.: 5.1e-156 Length: 743
Score: 155.00 Matches: 228
Percent Similarity: 98.28% Conservative: 0
Best Local Similarity: 98.28% Mismatches: 2
Query Match: 21.99% Indels: 4
DB: 5 Gaps: 0

US-10-028-384-12 (1-705) x BU471816 (1-743)

| | | | |
|----|-----|--|-----|
| Qy | 380 | SerAsnLeuSerAspAlaArgllePheIleMetTyrGlyValThrSerMetTyrPhe | 399 |
| Db | 5 | AGCAATCTCAGATGCCGCAATTTTATCATCATGTATGGCGTACCTCATGTACTTC | 64 |
| Qy | 400 | SerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG1 | 419 |
| Db | 65 | TCGTGCTGATGTCGGCTGATGCTGGTGCTGGCCCGGTGATGTG-CATCCTTTCCGG | 123 |
| Qy | 419 | YlleGlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProAs | 439 |
| Db | 124 | CATCGCGGTTTCAGGTGTGTTCACCTACATGAAGAACCTGGACATCAGCGGCCGAGA | 183 |
| Qy | 439 | pLysLysSerIysGlnGluAspSerThrTyrProIle-LysIleGluValAlaSerG | 459 |
| Db | 184 | CAAGAGAGCAAAAGACAGACACTCCACCTACCCCATCAAAAT-CAAGTTCGCCAGG | 242 |
| Qy | 459 | lytIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValT | 479 |
| Db | 243 | GCATGATCCCTGGGTGGTTTCTTCCTCATCAGTACACTTTCACCTCGACCTGGGTGA | 302 |
| Qy | 479 | hrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerA | 499 |
| Db | 303 | CCAGCGAGGCATACTCATCGCCCTCCATCGTGTGTGAGCGCGCGGCGGACGCGCAGCA | 362 |

| | | | |
|----|-----|---|-----|
| Qy | 499 | rgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluA | 519 |
| Db | 363 | GAATCATCTTTGATGACTTCAGGGAGGCGCTACTACTGCTGGGCACACACACCGAGG | 422 |
| Qy | 519 | spAlaIysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgT | 539 |
| Db | 423 | ACGCCAAGGTGATGTCCTGGTGGGACTACGGCTACGAGATCACCGCATGGCCCAACGCA | 482 |
| Qy | 539 | hrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaM | 559 |
| Db | 483 | CCATCTGCTGGTGACAAACACACCTGGAAACACACGACATCTCCCGTGTGGCAGCGCA | 542 |
| Qy | 559 | etAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValL | 579 |
| Db | 543 | TGGCATCGACGGAGGAGAAAGCTTACGAGATCATGAGAGAGCTGGACGTGAGTACGTGC | 602 |
| Qy | 579 | euValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheIleuTrpM | 599 |
| Db | 603 | TGGTGTATCTTCGGCGGCTCACCGGCTATTTCGTCGACGACATCAACAAAGTTCTCTGTGA | 662 |
| Qy | 599 | etValArgIleGlyGlySerThrAspThrGly | 609 |
| Db | 663 | TGGTGGCATCGCGGCGGACGACAGACACGCGGC | 694 |

Search completed: December 15, 2004, 06:14:40

Job time : 5478.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:51:45 ; Search time 7810.27 Seconds
(without alignments)
4268.640 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKGFRLSYEKQDTLLKL.....HMLVRIYKVVDLNRGLSRT 705

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131519_8017/app.query.fasta_1.3740
-DB=GenEmbl -QWMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd -LIST=45
-DOALIGN=200 -THR_SCORES=200 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn_1_1_20262@runat_14122004_131519_8017 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DGPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pri.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 3697 | 100.0 | 2472 | 6 | AX799092 Sequence |
| 2 | 3697 | 100.0 | 2472 | 9 | I38961 Human putat |
| 3 | 3657 | 98.9 | 2118 | 9 | BT007100 Homo sapi |
| 4 | 3657 | 98.9 | 2118 | 12 | BT008132 Synthetic |

RESULT 1
AX799092
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.:
Score:

AX799092
Sequence 11 from Patent WO03054008.
AX799092
AX799092.1 GI:37605063
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1
Perreault,C. and McBride,K.
Mammalian SIMP protein, gene sequence and uses thereof in cancer
therapy
Patent: WO 03054008-A 11 03-JUL-2003;
Compatisigne Inc. (CA)
location/Qualifiers
1. .2472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/note="ITM1, accession No NM-002219"

2472 bp mRNA linear PAT 08-OCT-2003

3657 98.9 2450 9 BC048348
3657 98.9 2458 6 BC034714
3657 98.9 2516 9 BC020965
3652 98.8 2472 6 CQ727867
3647 98.6 2283 9 HUMTMC
3644 98.6 2898 9 HSM809254
3633.5 98.3 2435 6 CQ834710
3633 98.3 2736 10 BC037612
3633 98.3 3094 6 AX799090
3633 98.3 3094 10 MSTRPR
3631 98.2 2760 6 CQ492624
3625 98.1 3666 10 AK129027
3598 97.3 2481 5 BC067313
3555 96.2 2567 5 BC063234
3555 96.2 2608 5 BC046072
3339 90.3 2329 6 CQ834708
2903.5 78.5 2855 6 CQ573675
2901 78.5 2470 6 CQ834712
2644 71.5 2339 6 CQ834706
2602.5 70.4 6153 6 CQ573674
2602.5 70.4 18405 2 AC017914
2602.5 70.4 16908 3 AC104631
2602.5 70.4 283556 3 AE003589
2268.5 61.4 2555 8 BT012881
29 60.8 2481 6 AX799082
30 2247 60.8 2481 9 AY074880
31 2237 60.5 2710 6 AX799084
32 2237 60.5 2756 8 AK099138
33 2235 60.5 4236 10 BC052433
34 2227 60.2 2819 8 AK067609
35 2214 59.9 1522 5 BC064436
36 2210 59.8 1421 5 BC043914
37 2177.5 58.9 2417 3 AF132552
38 2177.5 58.9 2417 6 AX799088
39 2177.5 58.9 2699 6 CQ589353
40 2103 56.9 4922 6 CQ589352
41 2103 56.9 162921 3 AC007853
42 2103 56.9 181132 3 AC008206
43 2103 56.9 227219 3 AE003750
44 2076 56.2 75650 2 AC018145
45 2070.5 56.0 2256 6 AX489023

ALIGNMENTS

Length: 2472
Matches: 705

Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 100.00%
 DB: 6

Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-10-028-384-12 (1-705) x AX799092 (1-2472)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 107 ATGACTAAGTTTGGATTTTGGATTTGCTATGAGAACGACACACTTTTGAAGCTT 166
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 167 CTCATTCTGCAATGGCTGCTGATTTATCTCTCCACCTGCTGTTGCTGCTCGAGA 226
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 DB 227 TTTGAAAGTGTATCCATGATGTTTGAATCCGTACTTTTAATTATCGGACTACAGGTTCTCTG 286
 QY 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrPheLeu 80
 DB 287 GCTGAGAGGGGTTTATAAATCCATAAAGCTGTTTGAAGCCGAGCCTGGTACCCCTTTG 346
 QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 347 GGACGAATCATGGAGGAACAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 407 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGTCTGTGTGTTCTCTGCCC 466
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 DB 467 CCT 526
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 DB 527 GCAGGGCTGGGCT 586
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 DB 587 TCTGTGCTGGCTCTCTATGATAATGAAGGATTTGCCATCTTTTGCATCTCACTCACTAC 646
 QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 DB 647 TACATGTGGATCAAGGACGAAAGACTGGTTCCATCTGTGGCAGCTAAGTGTGCTCTT 706
 QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuLeuAsnLeuIlePro 220
 DB 707 GCTTATTTCTACATGGTCTCGTCAATGGGAGGTTATGTCTCTCTCACTCACTTAATCTCT 766
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 767 CTCACAGTCTCTGTGCTGATGCTCACAGGCGGTTCTCTCACCGGATCTATGTGCTCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 827 TGTACTGTTTACTGCTGCTGGTACTATATCTTTTAGGCAGATCTCTCTGTGGTTCGAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 887 CCTGCTCTCTTATCAGAGACACATGGCAGGGTTTGGGGTCTTTGGTCTCTGCGACATCCAT 946
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 DB 947 GCCTTTTGTGATACCTGCGCAGCAAGTTCAATCCACAACTTTTGAAGTTCTTTTCCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyValLeuLeuMetLeuThr 320
 DB 1007 AGCGTCACTCTCTGTGAGCTTTGCTCTCTCACTGGGAGCTCTCTCTCATGCTGACA 1066
 QY 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340

DB 1067 GGAAAAATATCTCCCTGGACGGGGCGTTTCTACTCTACTGCTGGATCCCTCTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 DB 1127 AACAAATCCCATCATCTGCT 1186
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1187 TATTTTGGCTGACAGCT 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1247 AACCTGTCTGATGCCGAGTTTATCATCATGATGCTGTGACAGCATGACTTTTCTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1307 GCTGTAATGGTGGCTCTAATCTAGTGTGTGCACCTGTTATGAGCATTTCTCTCTGCAAT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 DB 1367 GGAGTCTCCAGGCTGCTGCTCCATCATCATGAGAACTTGGACATTAAGTCCGCCACAG 1426
 QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1427 AAGAGCAAGAAACACAGGATTCACCTACCTTAAAGATTGAAGTGGCAAGTGGGATG 1486
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1487 ATACGTGTCATGGCTTCTTCTCTCATCTACCTTCACTTCACTTCACTTCACTTCACT 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1547 GAGGCTACTTCT 1606
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1607 ATATTGTGATGACTCCGAGAAAGCAATATTATGCTTCTCTCTCTCTCTCTCTCTCT 1666
 QY 521 LysValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1667 AAGGTCACTGCTGCTGGATTTATGGCTATCAGATTACAGCTATGGCAAAACCAAT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1727 TTATGTGGACATACACATGGAATATATCCCATATTTCTCGAGTAGGAGGCAATGGG 1786
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1787 TCCACAGAGGAAAAAGCCATGAGATCATGAGGAGCTCGATGCTAGCTATGTGCTGGTC 1846
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 DB 1847 ATTTTGGAGGCTCTACCTGGGATTTCTCTGATGATATCAACAGTTTCTTTGGAAGTC 1906
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 DB 1907 CGGATTGGAGGAGGACAGATACAGGCAAAACATATCAAGGAGATGACTATTATATCTCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1967 ACTGGGAGTTCCGTTGGACCTGGAAGTTTCTCCAGTGTCTCACTGCTCATGATAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCCAGGCTCTCTCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 2087 GACCTGTCCGAATGCTGAGATTGGGAATAGAGACTTTCAGCTTGTATGCTCTGGAGGAA 2146
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 DB 2147 GGCTATACACAGAACATTTGCTGCTGAGGATATACAAAGGTAAAGGACCTGGGATAATCGA 2206

```

QY 701 GlyLeuSerArgThr 705
Db 2207 GCCTTGTCAGGACA 2221

RESULT 2
HUMBSA 2472 bp mRNA linear PRI 14-AUG-1996
LOCUS Human putative transmembrane protein precursor (B5) mRNA, complete
DEFINITION
ACCESSION L38961
VERSION L38961.1 GI:624703
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2472)
AUTHORS Lissy,N.A., Bellacosa,A., Sonoda,G., Miller,P.D., Jhanwar,S.C. and
Testa,J.R.
TITLE Isolation, characterization, and mapping to human chromosome
11q24-25 of a cDNA encoding a highly conserved putative
transmembrane protein, TMC
JOURNAL Biochim. Biophys. Acta 1306 (2-3), 137-141 (1996)
MEDLINE 96221283
PUBMED 8634329
FEATURES
source
    1..2472
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
gene
    1..2472
    /gene="B5"
5'UTR
    1..106
    /gene="B5"
CDS
    107..2224
    /gene="B5"
    /note="putative transmembrane protein precursor"
    /codon_start=1
    /protein_id="AAB05994.1"
    /db_xref="GI:624704"
    translation="MTKEFGFLRLSYEKQDTLLKLLLSMAAVLSPFSTRLPVLRPESEV
    LHEPDVENVETPEFLAEFGYKPHNWFDDRAVPLGRILIGGTIYVGLMITSAALVHV
    LHPHITIDRNVCVFLAPLESFSTISVITLLKELKADGALLAAMTAVPGFISR
    SVAGYNDEGAIFCMULTTYMMIKAVTGTCISICWAACALAFYVWSSGGVYFLLNL
    IPLVLMVTGRSHRIYVAYCTVCLTSRQISFVGFQVLSSEHMGFGVGL
    CQIHAFVDYLRSKLPQDFEVLRSVLSLVFLLVGLALLMTGKISDPAIFYSLL
    DPSYAKXNIPIIASVSRHQPTWSSVYFDQLLVFMPVGLVYCFENLSDARIFIMY
    GVTSMYSAMVRLMLVLPVMSILSGISVSVLTYMKNLDISRPDKSKKQOOSTY
    PIKIEVASGMILVMAFLITYTFTSTWVTSEAYSPSIVLSARGGDSRLIIDDFREA
    YXMRHNTPEDAKVMWDYGYQITAMARTILYDNTNNTNTHISRVGAMASTBEKA
    YEIMRELDVSVLVIQGLIGYSDDLNKLFWVRIGGSTDYGKHKEGNDYITPGF
    RVDREGSPVLLNCLMYKCYRFQGVYTEAKRPPGPDVRNABEIGNKDFELDVLEEGY
    TTEHLVRIYKYKDLNRLSRT"
sig_peptide
    107..196
    /gene="B5"
mat_peptide
    197..2221
    /gene="B5"
misc_feature
    452..459
    /product="putative transmembrane protein"
    /gene="B5"
misc_feature
    452..459
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    533..580
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    620..667
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    746..793
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    815..862
    /note="encodes putative transmembrane domain"
    /gene="B5"

```

```

misc_feature
    1013..1060
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    1199..1246
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    1310..1357
    /note="encodes putative transmembrane domain"
    /gene="B5"
misc_feature
    1475..1522
    /note="encodes putative transmembrane domain"
    /gene="B5"
3'UTR
    2223..2472
    /note="encodes putative transmembrane domain"
    /gene="B5"
polyA_site
    2455
    /gene="B5"
ORIGIN
Alignment Scores:
Pred. No.: 2,56e-311 Length: 2472
Score: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-028-384-12 (1-705) x HUMBSA (1-2472)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTTGGATTTTGGCATTTTGGCATTTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT 166
QY 21 LeuLeuLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCAATTCGTCATGGCTGCTGATTATCTCTCCACCTCGTCTGTTCTGCTGCTGAGA 226
QY 41 PheGluSerValLeuHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTCAGAGTGTATTCATGATGTTGATCCGTACTTTTAAATATCGGACTACCAAGGTTCTTG 286
QY 61 AlagluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaThrTyrProLeu 80
Db 287 GCTGAGAGGGGTTTATAAATTCATTAAGTGTGTTGATGACCGAGCTGGTACCCCTTG 346
QY 81 GlyArgLeuLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACCAATCATGAGGAGAACAAATTTACCCAGGTTTAATGATCACTCTCTGCTGCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGTCGTGTGTTCTTGCC 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526
QY 141 AlagluValAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 586
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 646
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGATCAAGGAGCAAGATGTTTCCATCTGTGGGACACTAAGTGTGCTCTCTCTCT 706
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuPro 220
Db 707 GCTTATTTCTACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 766

```

221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 767 CTCACGTCCTCGTGTGCTGATGCTCACAGGCGCTTCTCTCCCGGATCTATGTGGCTTAC 826
 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 827 TGTACTGTTTACTGCGCTGGGTACTATACTTCTTAGGCAGATCTCTCTTGTGGGTTCGAG 886
 261 ProValLeuSerSerGlnHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 887 CCGTCTCTTTCATCAGACACATGGCAGGGTTTGGGCTCTTGGTCTCTGCCAGATCCAT 946
 281 AlaPheValAspTyrLeuArgSerHisLeuAsnProGlnGlnPheGluValLeuPheArg 300
 947 GCCTTTGTGGATTACTCGCGCAGCAAGTTGAATCCACACAATTTGAAGTCTTTTCCGG 1006
 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuMetLeuThr 320
 1007 AGCGTCATCTCTGTGTAGGCTTTGTCTCTCTCACCGTGGAGCTCTCTCATGTGTACA 1066
 321 GlyIleSerProThrPheGlyArgPheTyrSerLeuLeuAspProSerTyrAlaIlys 340
 1067 GGAAATAATATCTCCCTGGAGCGGGCTTCTTCTACTGCTGGATCCCTCTTATGCTAAG 1126
 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 1127 AACCAACATCCCATCTGTCTGTGTGTGAGCATCAGCCACCAACCTGGTCTCTCATAC 1186
 361 TyrPheAspLeuGlnIleLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 1187 TATTTTACCTGAGCTCCGCTCTCATGTTCAGTTTTCAGTTGGCTCTATATCTGTCTAGC 1246
 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
 1247 AACCTGTCTGATGCCGGATTTTATCATCATGTATGTGTGACCCAGCATGTACTTTTCA 1306
 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyTle 420
 1307 GCTGTAAATGTGGGTCTAAATGCTAGTGTGGACCTGTGTATGAGCATTTCTCTGGCAT 1366
 421 GlyValSerGlnValLeuSerThrTyrMetIleAsnLeuAspIleSerArgProAspIlys 440
 1367 GGAGTCTCCAGGCTGTCTCCATCATCATGAGATCTCGACATTAAGTCGCCAGACAA 1426
 441 LysSerIlyslsGlnIleAspSerThrTyrProIleIleGluValAlaSerGlyMet 460
 1427 AAGAGCAAGAAAGCAACAGAGATCCACCTACCTATTAAGATTGAAGTGGCAAGTGGGATG 1486
 461 IleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 1487 ATACTGTGTATGCTTTCTTCTTCATCATCATCCTACCTACCTTTCATTAACCTGGTGACCA 1546
 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyValAspGlySerArgIle 500
 1547 GAGGCCCTACTTCTCTCGTCCATGTACTATCTGCCCGGTGGGGAGTGGCAGTAGGATC 1606
 501 IlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 1607 ATATTTGATGATCTCCGAGAAGCATATATTGGCTTCGTTCATTAATCTCCAGAGGATCG 1666
 521 LysValMetSerThrProPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 1667 AAGTCTATGCTCGGTGGGATTAAGCTATCAGATTACAGCTATGGCAACCGCAACAT 1726
 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 1727 TTAGTGACAATAACATGGAATAATACCCATATTTCTCGAGTAGGCGAGGCAATGGCG 1786
 561 SerThrGluGluIysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 1787 TCCACAGGAAAGAGCTATGAGATCATGAGGAGCTCGATGTGAGCTATGTCTGTGTC 1846
 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

1847 ATTTTGGAGGCTCCTCCTGGGTATTCCTCTGATGATATCAACAAGTTTCTTGGATGGTC 1906
 601 ArgIleGlyGlySerThrAspThrGlyHisIleIleIleGlyAsnAspTyrTyrThrPro 620
 1907 CGGATGGAGGAGCAGATACAGGCAACATATCAAGGAGATGACTATATATCTCA 1966
 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 1967 ACTGGGAGTTCCGTTGGACCGTGAAGGTTCTCCAGTGTCTCAACTGCCCTCATGTAC 2026
 641 LysMetCysTyrTyrArgPheGlyValTyrThrGluAlaIlysArgProGlyPhe 660
 2027 AAGATGTTTACTATCGCTTTGGACAGTTTACAGAGCCAGCGTCTCCAGGCTTT 2086
 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 2087 GACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCTCGGAGAA 2146
 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysVallyAspLeuAspAsnArg 700
 2147 GCTATATACACAGACATTTGCTGTGTCAGGATATACAGGTAAAGACCTGGATATATCGA 2206
 701 GlyLeuSerArgThr 705
 2207 GGCTGTGTCAGGACA 2221

RESULT 3
 BT007100 2118 bp mRNA linear PRI 13-MAY-2003
 LOCUS Homo sapiens integral membrane protein 1 mRNA, complete cds.
 DEFINITION
 ACCESSION BT007100
 VERSION BT007100.1 GI:30583038
 KEYWORDS FLI CDNA.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2118)
 Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
 Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
 Phelan,M. and Farmer,A.
 Cloning of human full-length CDSs in BD Creator(TM) System Donor
 vector
 Unpublished
 2 (bases 1 to 2118)
 Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
 Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
 Phelan,M. and Farmer,A.
 Direct Submission
 Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
 Circle, Palo Alto, CA 94303, USA
 This CDS clone is a part of a collection of human full length
 expression clones generated by BD Biosciences Clontech and the
 Harvard Institute of Proteomics. Each CDS has been cloned in two
 forms: with and without stop-codon (to allow fusion with C-terminal
 tag). The CDS has been directionally cloned using BD in-Fusion(TM)
 cloning system between the SalI and HindIII sites of the pDNR-DUAL
 vector. Additional sequences in the clone: 'ACC' after SalI site
 and before 'ATG' to provide Kozak consensus sequence; 'GG' after
 last codon and before HindIII site to maintain reading frame.
 Clone distribution: http://bioinfo.clontech.com/orfclones.
 Location/Qualifiers
 1..2118
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GH00707X1.0"
 /clone_lib="BD Creator(TM) CDS Library derived from MGC
 collection"
 /lab_host="DH5alpha T1 resistant"
 /note="Vector: pDNR-Dual"
 1..2118

FEATURES
 source
 1..2118
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GH00707X1.0"
 /clone_lib="BD Creator(TM) CDS Library derived from MGC
 collection"
 /lab_host="DH5alpha T1 resistant"
 /note="Vector: pDNR-Dual"
 1..2118

CDS

```

/codon_start=1
/product="integral membrane protein 1"
/protein_id="AAP35764.1"
/db_xref="GI:30583039"
/translation="MTKEGFLRLSEVEKQDTLLKLLLSMAAVLSFSTRFLFAVRFSEV
IHEPDVENVETTFLEAEFGYKPHNFDDRAWPLKRIIGGIIYFGIMTISAIIYHV
LHPHIIIDIRNVCVFLAPLPSSTIIIVTHLTKELDKAGALLAAAMIAVPAFYISR
SVAGYDNEGIAIFCMULTIYMMIKAVKGTISLQWAAKCALAYFVMVSSWGGYVFLINL
IPLHVLMLTRFSHR IYAYCTVYGLTISLMQISFVGFQVPLVSEHMAAFGVFL
CQIFAVDYLRKLNPOQFVLFPSVSLVGLVLLTVGALLMTGKISLPTWGRFYSLL
DPSYAKNNIPIIASVSHQPTWSSYVFDLQLVFMPVGLVYCFNSLSPARIFIMY
GVTSMYFSAVMRLMLVAPVWCILSGIVSOVLSTYMKNLDIRDCKSKKQDSTY
PIKNEVASGIMLWAFELITYTHSTWITSEYSSPVSILSARGGSGDRIIPDFREA
YYLHRNTPEDAKSNWDYGYQITANARTILVDNNTWNTNHSIRVGQAMASTBEKA
YEIMRELDVSLVLFQGLTGYSDDLNKFMLRWIRGGSTDTGKH KENDYYPTEGEF
RVDRGSPFLNCLMCKYVRFQGYVTEAKRPPGDFRVENABIGNKDFELDLVEAY
TTEHLVRIYKVKDLNRGLSRT"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 6 2e-308 Length: 2118
Score: 3657.00 Matches: 698
Percent Similarity: 99.15% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 6
Query Match: 98.92% Indels: 0
DB: 9 Gaps: 0

```

US-10-028-384-12 (1-705) x BT007100 (1-2118)

```

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 1 ATGACTAAGTTTGGATTTTTCGATTGCTCTATGAGAAGCAGACACACTTTTGAAGCTT 60

QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 61 CTATTCGTCAATGGCTGCTGATTATCTCTCCACTGCTGTTGCTGCTGCTGCTGCTG 120

QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 121 TTGMAAGTGTATCCATGATGTTTGTATCCGTACTTTAATATCGGACTACCAAGTTCCG 180

QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrProLeu 80
DB 181 GCTGAGAGGGGTTTATAAATCCATACTGTTGATGACCGAGCCTGGTACCCCTTGG 240

QY 81 GlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 241 GGACGATCATTCGAGGACAAATTTACCAAGTTTAAATGATCACCTCTCTGCAATCTAC 300

QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 301 CATGTACTCCATTTTTCACATCATCCATCGACATTCGGAATGTCGTGTGCTGCTGCTG 360

QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 361 CCTCTCTTCTCTCTCTCCACCATCGTCAGTACCCCTTACCAGAGAGCTCAAGGAT 420

QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
DB 421 GCAGGGGCTGGGCTTCTGCTGCTGCATGATTGCTGTAGTTCTCTGGATATATCTCCGA 480

QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 481 TGTGTGGCTGGCTCTCTATGATATGAAGGGATTGTCATCTTTTGGATGCTACTCACTTAC 540

QY 181 TyrMetTrrPheAlaValLysThrGlySerIleCysTrrPheAlaLysCysAlaLeu 200
DB 541 TACATGTGATCAAGCAGTAAAGACTGTTCCATCTGTTGGCAGCTAAGTGTGCCCTT 600

QY 201 AlaTyrPheTyrMetValSerSerTrrPheGlyTyrValPheLeuLeuAsnLeuPro 220
DB 601 GCTTATTCTACATGGCTCGTCATGGGAGGTTATGTGTTCTCTGATCAACTTAATCTCT 660

```

```

QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
DB 661 CTCACGTCCTCGTGTGATGTCTACAGGCCCTTTCTCTCACCGGATCTATGTGCGCTAC 720

QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
DB 721 TGTACTGTTTACTGCTGGCACTATATCTTTCTATGCAGATCTCTTTGTGGGTTTCCAG 780

QY 261 ProValLeuSerSerGluHisValAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
DB 781 CTTGCTCTTTATCATGAGACATGAGCCCTTTGGGCTCTTTGGTCTCTCTGCCAGATCCAT 840

QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
DB 841 GCCTTTTGGAATTACTCGCGCAGCAGTTGAATCCACAACTTTGAAGTCTTTTCCGG 900

QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
DB 901 AGCGTCATCTCTCGTAGGCTTTGCTCTCTCACCGTGGGAGCTCTCTCATGCTGACA 960

QY 321 GlyLysIleSerProTrrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
DB 961 GGAAAAATATCTCCCTGAGCGGGGCTTTCTACTCGCTGCGATGCCCTCTTATGCTAAG 1020

QY 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrrSerSerTyr 360
DB 1021 AACAAACATCCCATCATTTGCTGTCTGAGCATCAGCCACAACTGCTGCTCTCATAC 1080

QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
DB 1081 TATTTTGACCTGCGACTCTCTGCTTCAAGTTTCCAGTTGGCCTCTATTTCTGTTAGC 1140

QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
DB 1141 AACCTGTCTGATGCCGGATTTTATCATCATGATATGCTGTGACCATGATGATCTTTCA 1200

QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
DB 1201 GCTGTAAATGGTGGCTTAATGCTAGTGTGGCACCTGTTATGTGCATTTCTCTCTGGCAT 1260

QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
DB 1261 GGAGTCTCCAGGTGCTGTCACATACATGAGAAATCTGACATAGTCTGTCACAGCAAG 1320

QY 441 LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
DB 1321 AAGAGCAAGAGCAACAGGATTCACCTACCTTAAAGAAATGAAGTGGCAAGTGGCATG 1380

QY 461 IleLeuValMetAlaPhePheLeuLeuThrTyrThrPheHisSerThrTrrValThrSer 480
DB 1381 ATACTGGTCACTGGCTTTCTTTCTCATCACCTTACACCTTTTCTTCACTGGTGACAGT 1440

QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
DB 1441 GAGGCTACTCTTCTCGCTCCATTTGATCTATCTGCCGCTGGTGGGATGTCAGTAGGATC 1500

QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrrPheArgHisAsnThrProGluAspAla 520
DB 1501 ATATTGATGACTTCCGAGAGCATATATTATGGCTTCTGTCATATCTCCAGAGGATGCG 1560

QY 521 LysValMetSerTrrTrrPheAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
DB 1561 AAGGTCATGCTCTGGTGGGATTTATGGCTATCAGATTACAGCTATGGCAACCCAGACAA 1620

QY 541 LeuValAspAsnAsnThrTrrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
DB 1621 TTAGTGACAAATTAACACATGGAATAATACCCATATTCTTCTGAGTAGGAGGCAATGGCG 1680

QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
DB 1681 TCCACAGAGAAAAGCCATGAGATCATGAGGAGGCTCGATGCTAGCTATGTGCTGGTC 1740

QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrrMetVal 600

```

```

1741 ATTTTGGAGGCTCAGTGGGTATCTCTGATGATATCAACAGTTCTTTGGATGTC 1800
QY 601 ArgIleGlySerThrAspThrGlyHisIleGlyLeuAsnAspTyrThrPro 620
DB 1801 CGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGATATATATCTCA 1860
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
DB 1861 ACTGGGAGATTCCTGTGGACCGTGAAGTTCTCCAGTCTGCTCACTGCCCTCATGAC 1920
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysAspProGlyPhe 660
DB 1921 AAGATGTGTTACTATCGCTTTGGACAGGTTTACACAGAGCAAGCGTCTCCAGGCTTT 1980
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
DB 1981 GACGCTGTCCGAATGCTGAGATTGGGATTAAGACTTTGAGCTTGATGCTCTGGAGGA 2040
QY 681 GlyTyrThrThrGluHisTyrPheValArgIleTyrLysValLysAspLeuAspAsnArg 700
DB 2041 GCATATACACAGAACATTCGCTGTCAGGATATACAGGTAAAGGACCTCGATATCGA 2100
QY 701 GlyLeuSerArgThr 705
DB 2101 GCCTTGTCAAGGACA 2115

RESULT 4
LOCUS BT008132
DEFINITION Synthetic construct Homo sapiens integral membrane protein 1 mRNA,
partial cds.
ACCESSION BT008132
VERSION BT008132.1
KEYWORDS GI:30585102
SOURCE synthetic construct
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2118)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2118)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
source
1..2118
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone_lib="BD Creator(TM) CDS Library derived from MOC
collection"
/lab_host="DH5alpha T1 resistant"
/notes="Vector: pDNR-Dual"
1..>2118
CDS

```

```

/notes="Mutations: 2117:Stop->Leu"
/codon_start=1
/transl_table=1
/product="Homo sapiens integral membrane protein 1"
/protein_id="AAP36824.1"
/db_xref="GI:30585103"
/translation="MTKFGFLRLSYEKQDTLLKLLLSMAAVLSFSTRFLPAVRFESV
LHEPDYNYTTRFLAEEGYPKHNFDRAWYPLGRIIGGTIYPLGMITSAIYHV
LHPHTIDIDNVCEVLAPLSESTTIVYHLKELKADAGALLAAIAVPGYLSR
SVAGSYDEGIAIFCMULTYNNKAVTSGICWAACALAYFWVSSKGGYVFLNL
LPLHVLMLFRFSHRIYVAICTYICITLSMOISFVGFQVLSERMAAFGVGL
COIHAVDYLRSKLNPOQFVLSVLFVLLTVGALMLTGKISPMTPGRFYL
DPSYAKNNPIIASVSEHQPTTWSSYFDLQLLFVFPVGLYFCFNLSARFIILY
GTYMYFGAVMRLMLVLPWCILSGIVSOVLSTYMKWLDISRPDKSKKQDDTY
PKNEVASGMILVNAFLITVYFHSITWVSEAYSSPSIVLSARGGDSRIIFDPPEA
YVLRHNTPEDAKMSWMDYQITAMANTILVDNNTIIVDNTNTHISVAGQMASTBKA
YEIMRELDVSVLVIFGLTGYSSDDINKFLMWVRLGGSTDTGCKHKENDYTPTEGF
RYDREGSVLLNCLMYKMYRFGQVYTEAKRPPGPDVRVNAEIGNKDFELDLVLEAY
TTEHMLVRIYKQDLNDRGUSRII"
Alignment Scores:
Pred. No.: 6,2e-308 Length: 2118
Score: 3657.00 Matches: 698
Percent Similarity: 99.15% Conservative: 1
Best Local Similarity: 99.01% Mismatches: 6
Query Match: 98.92% Indels: 0
DB: 12 Gaps: 0
US-10-028-384-12 (1-705) x BT008132 (1-2118)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
DB 1 ATGACTAAGTTGGATTTTTCGATTTGCTTATCGAGAGGAGGACACACCTTTTGAAGCTT 60
QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 61 CTCATTCTGCTCAATGGCTGCTGATTATCTCTCCACCTGCTGTTTGTCTGCTCTGAGA 120
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 121 TTGGAAGTTGTTATCCATGAGTTGATCCGACTTTAATATCGGACTACACGGTTCCTG 180
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaIleTyrProLeu 80
DB 181 GCTGAGGAGGGGTTTATAAATTCATAACTGTTTGTATGACCGAGCTGCTGCTACCTTG 240
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 241 GGACGAAATCATGGAGGAACTTTACCCAGTTTATGATCATCCTCTGCTGCAATCTAC 300
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 301 CATGTACTCCATTTTTCACATCCATCCGATTCGGAATTCGTAATGCTGTGTGTCTCTGGCC 360
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 361 CTCCTCTCTCTCTCTCTCCACCATCGTCACTACCATCTTACCAAGAGCTCAAGGAT 420
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
DB 421 GCAGGGGCTGGGCTCTCTTCTGCTGCCATGATTCCTGTAGTTCTCTGATATATCTCCGA 480
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 481 TCTGTGGCTGGCTCTCTATGATTAATGAGGAGATTCCTCTTTGCTGCTACTACCTAC 540
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
DB 541 TACATGTGATCAAGGAGTAAAGACTGGTTCCTCTGTTGGGAGCAAGTAAAGTGTGCCCT 600
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuLeuLeuLeuPro 220

```


Db 601 GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTCTCTGATCAACTTAATTCCT 660
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 661 CTCACAGTCTCTGCTGCTGATGCTCACAGGCGTTTCTCTCACGGATCTATGTGGGCTAC 720
 Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 721 TGTACTGTATTACTGCTGGGCACTATATCTTCTATGCAGATCTCTCTTGTGGGTTTCCAG 780
 Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 781 CTTGTCCTTTTATCAGACACATAGGACCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 840
 Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 841 GCCTTTGTGGATTACTCTGGCAGCAAGTTGAATCCACAACTTTGAAGTCTCTTTTCCCG 900
 Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 901 AGCGTCATCTCTGGTAGGCTTTGCTCTCTCCAGTGGAGCTCTCTCTCATGCTGACA 960
 Qy 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 961 GGAATAATATCTCCCTGGACGGGGCTTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1020
 Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrSerSerTyr 360
 Db 1021 AACACATCCCATCATGCTCTCTCTGTCGAGCATCAGCCACACCTGGTCTCTCATAC 1080
 Qy 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1081 TATTTTGCCTGACGCTCTCTGCTCTCTCATGTGTCAGTGGCTCTATTAATCTGCTTAGC 1140
 Qy 381 AsnLeuSerAspAlaArgPheIleMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1141 AACCTGTCTGATGCCGGAATTTTATCATCATGATGCTGACCATCATGACTTTTCA 1200
 Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1201 GCTGTAAATGGTGGCTTAATGCTAGTGTGGACGCTGTATGTGTGATCTCTCTGGCAT 1260
 Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1261 GGAGTCTCCAGTCTCTGTCACATCATCATGATGATGATGATGATGATGATGATGATGAT 1320
 Qy 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1321 AAGAGCAAGCAAGCAAGCAAGTCCACCTACCTTAAAGATGAAGTGGCAAGTGGGATG 1380
 Qy 461 IleLeuValMetAlaPhePheIleThrTyrThrPheHisSerThrTyrValThrSer 480
 Db 1381 ATACTGGTATGGCTTTCTTTCTCATCATCATCATCATCATCATCATCATCATCATCATCAT 1440
 Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1441 GAGGCTACTCTCTCCGCTCATCTACTATCTGCGCGTGGTGGGATGCGAGTAGGATC 1500
 Qy 501 IlePheAspPheArgGluAlaTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 520
 Db 1501 ATATTGTAGTATCCGAGAGCATATTATTGGTCTGTCATATATCTCAGAGGATGGC 1560
 Qy 521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1561 AAGGTATGCTCTGGTGGATTTATGGCTATCATCATCATCATCATCATCATCATCATCATCAT 1620
 Qy 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1621 TTATGGGCAATAACACATGGATTAATACCATATTTCTCAGTAGGCGAGGCAATGGCG 1680
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1681 TCCACAGAGCAAGCAAGCAAGTATGAGATCATGAGGAGCTCGATGTGCTATGTGCTGGTC 1740

Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetVal 600
 Db 1741 ATTTTTCGAGCCTCACTGGGTATCTCTGATGATATCAACAAGTTTCTTTGGATGGTC 1800
 Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1801 CGGATTGGAGGGAGCAGATACAGCAACATATCAAGGAGATGACTATTATTAATCTCA 1860
 Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1861 ACTGGGAGTTCCGTGGACCGTGAAGGTTCTCCAGTCTGCTCACTCACTCATGTAC 1920
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 1921 AAGATGTGTACTATCTCTTGGACAGGTTTACAGAGCCAGGCTCTCCAGGCTTT 1980
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 1981 GACCGTCTCCGAAATCTGAGATTGGGAATAAAGACTTTGAGCTTCTGCTGGAGGAA 2040
 Qy 681 GlyTyrThrThrGluHisTyrProLeuValArgIleTyrLysValLysAspLeuAspLeuArg 700
 Db 2041 GCATATACACAGAACATTTGGCTGGTTCAGGATATACAGTAAGGACCTGGATAATCGA 2100
 Qy 701 GlyLeuSerArgThr 705
 Db 2101 GGCTTGTCAAGGACA 2115

RESULT 5

BC048348

LOCUS

DEFINITION

IMAGE:5561615, partial cds.

ACCESSION

BC048348

VERSION

BC048348.1

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 2450)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,

Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Faney, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,

Burtchfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,

Sutcher, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

REFERENCE

2 (bases 1 to 2450)

AUTHORS

Strausberg, R.

Direct Submission

Submitted (07-MAR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK

COMMENT

Contact: MGC help desk

BC048348 2450 bp mRNA linear PRI 25-JUN-2004
 Homo sapiens integral membrane protein 1, mRNA (cdna clone
 IMAGE:5561615), partial cds.
 BC048348
 BC048348.1 GI:29477243
 Homo sapiens (human)

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2450)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Faney, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Burtchfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,
 Sutcher, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932

2 (bases 1 to 2450)

Strausberg, R.

Direct Submission

Submitted (07-MAR-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

| | | | |
|----|------|---|------|
| Db | 46 | ATGACTAAAGTTGGATTTTGGGATTTGCTATGAGAGCAGGACACACTTTTGAAGCTT | 105 |
| Qy | 21 | Leu1LeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg | 40 |
| Db | 106 | CTCATCTCTCAATGGCTGCTGATTAATTCCTTCCACTCTGCTGTTGCTGCTGAGA | 165 |
| Qy | 41 | PheGluSerValLeuHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu | 60 |
| Db | 166 | TTTGAAGTGTTATCATGAGTTTGATCCGTTTAAATATCGAGTACCAGGTTCCTG | 225 |
| Qy | 61 | AlaGluGluGlyPheTyrIlePheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu | 80 |
| Db | 226 | GCTGAGGAGGGGTTTATAAATCCATAACTGGTTTGAAGACCGAGCTGTGTACCTTTG | 285 |
| Qy | 81 | GlyArgIleLeuGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr | 100 |
| Db | 286 | GGACGATCATTTGGAGAACATTTACCCAGGTTTATGATCACTCTGCTGCAATCTAC | 345 |
| Qy | 101 | HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla | 120 |
| Db | 346 | CATGTACTCCATTTTTCACATCCACATCGCATTCGGAATGCTGTGTCTCTGCGCC | 405 |
| Qy | 121 | ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp | 140 |
| Db | 406 | CCTCTCTTCTCTCTCTTACCACCATCGTCAGTACACCTTACCAAGAGCTCAAGAT | 465 |
| Qy | 141 | AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg | 160 |
| Db | 466 | GCAGGGGCTGGGCTTCTGCTGCTGCATGATTTGTGTAGTTCTGTGATATATCTCCGA | 525 |
| Qy | 161 | SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr | 180 |
| Db | 526 | TCTGTGCTGCTGCTCTTATGATTAATGAAGGATGTCATCTTTTGGATGCTACTACCTAC | 585 |
| Qy | 181 | TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu | 200 |
| Db | 586 | TACATGTGGATCAAGGCAGTAAAGCTGTTCCATCTGTGTGGCAGTAAAGTGTGCCCTT | 645 |
| Qy | 201 | AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro | 220 |
| Db | 646 | GCTTATTTCTACATGCTCTCGTCATGGGAGGTTATGTTCTCTGATCAACTTAATCTCT | 705 |
| Qy | 221 | LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr | 240 |
| Db | 706 | CTCAAGTCTCTGCTGCTGATGCTCAGAGGCGTTTCTCTCACGGATCTATGTGGCTAC | 765 |
| Qy | 241 | CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln | 260 |
| Db | 766 | TGTACTGTTTACCTGCTGGGCACTATATCTTCTATGTCAGATCTCTCTTGTGGGTTCAG | 825 |
| Qy | 261 | ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis | 280 |
| Db | 826 | CCTGTCTCTTTCATCAGAGCAGTGGGAGGCTTTGGGCTCTTTGGTCTCTGCCAGTCCAT | 885 |
| Qy | 281 | AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg | 300 |
| Db | 886 | GCCTTTCTGGATTACCTGCGCAGCAAGTTGAATCCACAATTTTGAAGTCTCTTTCCGG | 945 |
| Qy | 301 | SerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetLeuThr | 320 |
| Db | 946 | AGGCTCATCTCTCTGGTAGGCTTTGCTTCTCCTCCTGCGGAGCTCTCTCATGCTGACA | 1005 |
| Qy | 321 | GlyLysIleSerProTyrThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys | 340 |
| Db | 1006 | GGAAATAATCTCCCTGAGCGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGTGAAG | 1065 |
| Qy | 341 | AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr | 360 |
| Db | 1066 | AACAACATCCCATCATTTGCTTCTGTGCTGAGCATCAGCCCAACACCTGGTCTCTATAC | 1125 |
| Qy | 361 | TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer | 380 |
| Db | 1126 | TATTTGACCTGAGCTCTCTCGTCTTCTATGTTTCCAGTGTGGCTCTATTACTGCTTAGC | 1185 |

Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dierich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-I., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 93 Row: P Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 22749414.

FEATURES
Location/Qualifiers
1..2450
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5561615"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH MGC_88"
/lab_host="DH103"
/note="Vector: pCMV-SPORT6"
1..2450
/gene="ITM1"
/note="synonyms: TMC, MGC9042"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
1..2163
/gene="ITM1"
/codon_start=1
/product="ITM1 protein"
/protein_id="AAH48348.1"
/db_xref="GI:29477244"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
/translation="SRPADRVLPPIHVQTKFGLRLSYEKQFLRLSKLILSMAAVL
SPTRFLAVLRSEVTHFEDPVYRTRTFLEAEFGYKFNWDFDRAWYPLGRIGGT
IYPLMTSAAIYVHLFHIIIDINVCVFLAPLESFTIIVYHLTKELKADAGL
LAAAMTAVFGYISRSVAGSDYNEGIAIFCMLLTIYMIKAVKTSISWAAKCALAYF
YVSSWGGYVFLINILPVLVMTGRFSRIYVAYCTVYCLGILNSQISFVGFQP
VLSEHMAAFVFGLCQIHAFVDYLRSLNPOQFVLFPSVISLVGFLLTVGALLML
TGKISPTGKGFSLDPSYAKNNIPIIASVSEHQPTSSYVFDLQILVFMFVGLYI
CFNSLSDAIFILMVGVTSMYFSAMVRLMLVLPWCILSGIGVSOVLSTYMKNDI
SRPKSDSKQSDSTPIKNEVASGMLVAFFLIYTHSTWVTSEAVSSPSIVLSAR
GGSKRIIFDREAYNLRHNPEDAKVMSWDYGYOITAMANEITLVDNNTWNTH
ISRVGQAMASTEKEIEMRELDVSIVLFGILGYGSSDDINKFLMWVRIGGSTDTG
KHINKENDYPTTFEFDVREGFVLLNLMYKNCYIRFGQVTEAKRPPGDFVRNAE
IGNKDFELDLAEAYTTEHVLVRIYKVDLNRGLSRT"

| | | | |
|------------------------|-----------|---------------|------|
| Alignment Scores: | 7.59e-308 | Length: | 2450 |
| Pred. No.: | 3657.00 | Matches: | 698 |
| Score: | 99.15% | Conservative: | 1 |
| Percent Similarity: | 99.01% | Mismatches: | 6 |
| Best Local Similarity: | 98.92% | Indels: | 0 |
| Query Match: | 9 | Gaps: | 0 |

DB: 9

US-10-028-384-12 (1-705) x BC048348 (1-2450)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20

Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2516)
 Strausberg, R.
 Direct Submission
 Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 25, 2003 this sequence version replaced gi:18089244.
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue procurement: DCTD/BTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 14 Row: e Column: 22
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22749414.

FEATURES

source

gene

1..2516
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:9042 IMAGE:3891543"
 /tissue_type="Lung, carcinoma, large cell undifferentiated."
 /clone_lib="NIH MGC_69"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

CDS

1..2516
 /gene="ITMI"
 /note="synonyms: TMC, MGC9042"
 /db_xref="LocusID:3703"
 /db_xref="MIM:601134"
 93..2210
 /gene="ITMI"
 /codon_start=1
 /product="integral membrane protein 1"
 /protein_id="AAH20965.1"
 /db_xref="GI:18089245"
 /db_xref="LocusID:3703"
 /db_xref="MIM:601134"
 /translation="MTKFGFLRLSYEKQDTLLKLLLSMAVLFSFTRFLAVLRFSV
 IHEFDYFNATTFLEAFEGYKPHNFDDRWYPLGRIICGTIYGLMITSAAIYHV
 LHFPHITIDNVCVFLAPLFSFTTIVTTLTKELKADAGALLAAMIAVPGYTSR
 SVAGSDNEGALICMLTYWIKAVKGTGICWAAKCALAYFWMVSGWGVFLNL
 IPLAVLMLTGRSHRIYVACTVYCLGILSLNQISFVGFQVLSSEHMAFGVFL
 CQIFADYLRKSNPQFELFRFSLVSLVGLVLLTVGALLMTGKISLWTRFVSL
 DPFSAKNPIIASVSHQPTWSSYFDLIQVFPFVGLYFCFNLSDARFIIMY
 GVTNYPFAMVRLMLVLPVCLISGIVSVLSTYMKLDSRPDKSKKQDSTY
 PIKQVAGSMILNVAFLITITFSTWTSSEYSSPILSARGGSGSRILIFDDFREA
 YWLRHNTPEAKVSWMDYGYQITAVANTILVDNNTWNTHTISRVGMASTESKA
 YVRELDVSVYLVIFGLTGYSDSDINKFLMVRIGGSTDTGKHKENDYTPTEGF
 RVDREGSVLANCLWYKVCYRFQVYTEAKRPGFDRVFNALGNKDFELDLVLEAY
 TTEHVLVYIKVLDNRGLSRT"

ORIGIN

Alignment Scores:

Pred. No.: 7.86e-308 Length: 2516
 Score: 3657.00 Matches: 698
 Percent Similarity: 99.15% Conservative: 1
 Best Local Similarity: 99.01% Mismatches: 6
 Query Match: 98.92% Indels: 0
 DB: 9 Gaps: 0
 US-10-028-384-12 (1-705) x BC020965 (1-2516)
 QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
 Db ATGACTAAGTTTGGATTTTTCGATGTCCTATGAGAAGCAGACACACTTTTGAAGCTT 152
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaLeuArg 40
 Db 153 CTCATTCTGTCATAGGCTGCTGATTATCTCTTCCACTCGTCTGTGTGCTGCTCGAGA 212
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 Db 213 TTTGAAAGTGTATCCATGAGTTTGATCCGCTACTTTAATATATCGGACTACCAAGTTCTCTG 272
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrProLeu 80
 Db 273 GCTGAGGAGGGGTTTATAAATCCATACTGGTTTGATGACCGACCTGGTACCTTTTG 332
 QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 333 GGACGAATCAITGAGGAAACAATTTACCAGGTTTAATGATCACCTCTGCTGCAATCTAC 392
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 393 CATGTACTCCAAATTTTCCACATCACCATCGAATTCGGAATGCTGTGTGTTCTCTGGCC 452
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 453 CCTCTCTTCTCCTCCTTACCACCATCGTCACGTACCATCTTACCAAGAGCTCAAGGAT 512
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 513 GCAGGGGCTGGGCTTCTGTCTGCCATGATGCTAGTCTGCTGATATATCTCCCGA 572
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 573 TCTGTGGCTGGCTCCTATGATAATGAAGGATTTGCCATCTTTTGCATGCTACTCACCTAC 632
 QY 181 TyrMetThrIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaLeu 200
 Db 633 TACATGTGATCAAGGCAGTAAAGACTGGTTCCATCTGTGTGGCAGCTAAGTGTGCCCTT 692
 QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
 Db 693 GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTTCTCGATCAACTTAATTCCT 752
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 753 CTCACGCTCCGCTGCTGATGCTCACAGGCCGTTCTCTCACCGGATCTATGTGGCCTAC 812
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 813 TGTACTGTTTACTCCCTGGSCACTATATCTTCTATGCGAGATCTCTTTTGGGGTTCCTAG 872
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheCysGlnIleHis 280
 Db 873 CCGTGTCTTTCATCAGAGCAGATGGCGCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 932
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 933 GCCTTTGTGGATTACCTGGCAGCAAGTTGAATCCACAAATTTGAAGTTCCTTTTCCGG 992
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 993 AGCGTCATCTCTCTGTAGGCTTTGTCTTCTCACCGTGGGAGCTCTCTCTCATGCTGAGA 1052

QY 321 GlyLysIleSerProThrThrGlyArgPheTyrSerLeuLeuAspSerTyrAlaLys 340
 DB 1053 GGAAAAATATCTCCCTGGAGCGGGCTTTCTACTCGCTCGTGAATCCCTCTATGCTAAG 1112
 QY 341 AsnAsnIleProIleLeuAlaSerValSerGluHisGlnProThrThrTrpSerTyr 360
 DB 1113 AACAAATCCCATCATTTCTGTGTCTGAGCATCAGCCACAACTGTCCTCATAC 1172
 QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1173 TATTTTGACCTGACAGCTCCCTGCTTCATGTTTCCAGTTGGCTCTATATCTGCTTTAGC 1232
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1233 AACCTGTCTGTATGCCCGATTTTATCATCATGTATGTTGTGACCAAGCATGTACTTTCA 1292
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1293 GCTGTATGTCGCTTAATGCTAGTGTGGACCTGTTATGTGCAATCTCTCTGGCAAT 1352
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 DB 1353 GGAGTCTCCAGAGTCTGTCCACATACATGAAGAACTCGGACATAGTCTCCAGACAAG 1412
 QY 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 DB 1413 AAGAGCAAGAGCAACAGGATTCACCTACCTATTAAAGTAAGTGGCAAGTGGATG 1472
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1473 ATACTGTGTCATGCTTCTCTCTCATCACCTACCTCATCACTGCTGGTGACCACT 1532
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1532 GAGGCTTACTCTTCTCGTCCATGTACTATCTGCCCCGTTGGGGATGGCAGTAGGATC 1592
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1593 ATATTTGATGACITCCGAGAAGCATATATTGCTTCGTCAATAATACTCCAGAGGATGCG 1652
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1653 AAGTCTATGCTTGGTGGATTTAGCTATACAGATTACAGCTATGGCAATCCCAACAAT 1712
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1713 TTAGTGACAATAACACATGGAATAATACCATATTTCTCGAGTAGGGCAGGCAATGGCG 1772
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1773 TCCACAGAGGAAAGAGCTATAGATCATGAGGAGCTCGATGTGATGTGTGCTGTC 1832
 QY 581 IlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 DB 1832 ATTTTGGAGGCTCTACTGGGTATCTCTGTATGATATCAACAAGTTCTTTGGATGTC 1892
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 DB 1892 CGGATTTGGAGGAGCACAGATACAGGCAACAATATCAAGAGAAATGACTATTTACTCCA 1952
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1952 ACTGGGAGTTCCGTGTGGACCGTGAAGTTCTCCAGTGTCTGCTCAACTGCTCATGTAC 2012
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
 DB 2012 AAGATGTGTACTATCGCTTTGGACAGTTTACACAGAGAGCCAGGCTCTCCAGGCTTT 2072
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 2072 GACCGTCTCCGAATGCTGAGATTGGGAATTAAGACTTTGAGCTTGATGTCTCTGGAGAA 2132
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700

DB 2133 GCATATACACAGACATTTGGTGTGAGGATATACAGGTAAGGACCTGGATATCGA 2192
 QY 701 GlyLeuSerArgThr 705
 DB 2193 GGCTTGTCAAGGACA 2207
 RESULT 8
 LOCUS CQ27867 2472 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 13801 from Patent WO02068579.
 ACCESSION CQ27867
 VERSION CQ27867.1 GI:42295093
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, B.W.
 TITLE Kits, such as nucleic acid arrays, comprising a majority of
 humanexons or transcripts, for detecting expression and other uses
 thereof
 JOURNAL Patent: WO 02068579-A 13801 06-SRP-2002;
 PE Corporation (NY) (US)
 FEATURES
 source
 1..2472
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,09e-307 Length: 2472
 Score: 3652.00 Matches: 697
 Percent Similarity: 99.15% Conservatives: 1
 Best Local Similarity: 99.01% Mismatches: 6
 Query Match: 98.78% Indels: 0
 DB: 6 Gaps: 0
 US-10-028-384-12 (1-705) x CQ27867 (1-2472)
 QY 2 ThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrIleLeuLysLeuLeu 21
 DB 110 ACTAAGTTTGGATTTTGGGATTTTCCATGAGAGGAGGACACATTTTGAAGCTTCTC 169
 QY 22 IleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPhe 41
 DB 170 ATTCTGTCAATGGCTGCTGTATTATCTCTCCACTCGTCTGTTGCTGCTGAGATTT 229
 QY 42 GluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAla 61
 DB 230 GAAAGTTTATCCATGAGTTTATCCCGTACCTTAAATATCGACTACAGGTTCTTGGCT 289
 QY 62 GluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeuGly 81
 DB 290 GAGGAGGGGTTTATAAATTCATACTGTTTGTATGATCCGAGGCTGTTACCTTTGGGA 349
 QY 82 ArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyrHis 101
 DB 350 CGAATCATTTGGAGGAACAATTTACCAGTTTAAATGATCACTCTGCTGCAATCTACCAT 409
 QY 102 ValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaPro 121
 DB 410 GTACTCCATTTTTCACATCACCATCGACATTCGAATGCTGTGTGTTCTTGGCCCT 469
 QY 122 LeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAla 141
 DB 470 CTCCTCTCTCTCTCTCCACCATGCTCACGTACCCTTACCAGAGCTTACAGATGCA 529
 QY 142 GlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArgSer 161
 DB 530 GGGGCTGGGCTTCTTGTCTGCTGCCATGATGCTGTAGTTCTGTGATATATCTCCCATCT 589


```
QY 162 ValAlaGlySerTyrAspAsnGluGlyLeuAlaIlePheCysMetLeuLeuThrTyrTyr 181
Db 590 GTGGCTGGCTCTATGATTAATGAAGAGGATGGCCATCTTTGGCATGTCTACTACCTACTAC 649
QY 182 MetTppileLysAlaValIleThrGlySerIleCysTyrAlaAlaLysCysAlaLeuAla 201
Db 650 ATGTGGATCAAGGAGTAAAGACGGTTCCATCTGTTGGCAGCTAAGTGTGCCCTTGCT 709
QY 202 TyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuLeuLeuLeuProLeu 221
Db 710 TATTTCTACATGGCTCTGCTCATGGGAGGTTATGTCTCTGATCAACTTAATTCCTCTC 769
QY 222 HisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCys 241
Db 770 CACGTCCTCGTGTGATGCTCACAGGCGCTTCTCTCAACGGATCTATGTGGCTACTGT 829
QY 242 ThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnPro 261
Db 830 ACTGTTTACTGCTGGGACATATCTTTCTATGAGATCTCCCTTGGGCTTTTCCAGCCT 889
QY 262 ValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAla 281
Db 890 GTCCCTTTCATCAGAGCACATGGCAGCCTTTGGGCTCTTGGTCTCTGCCAGATCCATGCC 949
QY 282 PheValAspTyrLeuArgSerIleLysLeuAsnProGlnGlnPheGluValLeuPheArgSer 301
Db 950 TTGTGGATTACCTCGCAGCAAGTGAATCCACAAATTTGAAGTCTTTTCCGGAGC 1009
QY 302 ValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGly 321
Db 1010 GTCATCTCTCTGGTAGGCTTTGTCTCTCACCGTGGGAGCTCTCTCTCATGCTGACAGGA 1069
QY 322 LysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsn 341
Db 1070 AAAATATCTCTCGACGGGCGTCTTCTACTGCTGGATCCCTCTTATGCTAGTAAGAC 1129
QY 342 AsnIleProIleAlaSerValSerGluHisGlnProThrThrTyrPheSerTyrTyr 361
Db 1130 AACATCCCATCATTTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1189
QY 362 PheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsn 381
Db 1190 TTGTACCTCGACGCTCTCTGCTTCTCATGTTTCCAGTGGCTCTTATCTCTTTAGCAAC 1249
QY 382 LeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSerAla 401
Db 1250 CTGTCTGATGCCGATTTTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1309
QY 402 ValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGly 421
Db 1310 GTAATGGTGGCTTAATGCTAGTGTGGACCTGTATGTGATCTCTCTCTGGCATTTGGA 1369
QY 422 ValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLys 441
Db 1370 GTCTCCAGGCTCTGCCACATACATGAAGAATCTGGACATAGTCTGCCAGACAAGAG 1429
QY 442 SerLysLysGlnIleAspSerThrTyrProIleLysIleGluValAlaSerGlyMetIle 461
Db 1430 AGCAAGAGACACACAGGATCCACCTTACCTTATTAAGAAATGAAGTGGCAAGTGGGATGA 1489
QY 462 LeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTyrValThrSerGlu 481
Db 1490 CTGGTCATGGCTTTCTTCTCATCATCATCATCATCATCATCATCATCATCATCATCAT 1549
QY 482 AlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIle 501
Db 1550 GCTACTCTCTCTCCGTCATGATGATCTATCTCCCGCTGGTGGGATGGCAGTAGATCATA 1609
QY 502 PheAspAspPheArgGluAlaTyrTyrTyrPheLeuArgHisAsnThrProGluAspAlaLys 521
Db 1610 TTGTATGATCTCCGAGAGGACATATTATGGCTTCTGCTCATATATCTCCAGGAGTGGAG 1669
```

522 ValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
1670 GTCAATGCTGGTGGGATTAATGCTATACAGATTACAGCTATATGCAAAACGACAACTTTTA 1729
542 ValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
1730 GTGGACAATAACACATGGAATAATACCATATTTCTCGAGTAGGCGAGCAATGGCGTCC 1789
562 ThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIle 581
1790 ACAGAGAAAGAGCTATGAGTCAATGAGGAGCTGATGTACCTATGTGTGTCAT 1849
582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArg 601
1850 TTTGAGGCTCTCACTGGGTATCTCTGATGATATCAACAAGTTCTTTGGATGGTCCGG 1909
602 IleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThr 621
1910 ATTGAGGGAGCACAGATACAGCAACATATCAGAGAGATGACTATTTATCTCCACT 1969
622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
1970 GGGGAGTTCCGTGTGGACCGTGAAGGTTCTCCAGTGTCTCAACTGCTCATGTACAAAG 2029
642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
2030 ATGTTTACTATCTCTTGGACAGGTTTACAGAGCCAGGCTCTCTCCAGCTTTGAC 2089
662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGly 681
2090 CCGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATCTCTGGAGGAGCA 2149
682 TyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArgGly 701
2150 TATACCACAGACATTTGGCTGTGCAGATATACAGTAAGACCTTGATATATCGAGGC 2209
702 LeuSerArgThr 705
2210 TTGTCAAGGACA 2221

RESULT 9
HUNTMC
LOCUS HUNTMC 2293 bp mRNA linear PRI 13-FEB-2002
DEFINITION Homo sapiens transmembrane protein (TMC) mRNA, complete cds.
ACCESSION L47337.1
VERSION L47337.1 GI:18654193
KEYWORDS transmembrane protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2293)
AUTHORS Bellacosa, A.A.B. and Testa, J.J.R.
TITLE Isolation, characterization and mapping to human chromosome 11q24-25 of a gene, TMC, encoding a highly conserved putative transmembrane protein
JOURNAL Unpublished
FEATURES
Location/Qualifiers
1..2293
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="11q24-25"
/cell_type="mesothelioma"
/clone_lib="ddc11"
1..2293
/gene="TMC"
117..2234
/gene="TMC"
/note="putative"
/codon_start=1
/product="transmembrane protein"
/protein_id="AAL77539.1"


```

/db_xref="GI:18654194"
/translation="MTKPGFURLSEYQDPTLLKLLLSMAAVLSFSTRLPVLRPSV
IHEPDYNYETTRFLBEEGYKFNWFDORAWPLGRIIGGTIYGLMITSAAIHSR
LHFPHITDIDNVCVFLAPFSFTTIVYHLTKELKADAGALLAAAMAVPGYISR
SVAGSDNEGIAICMLLYYMTIKAVTGSICWAAKCALAYFYVSSWGYYFLNL
IPLHVLMLGRSPSHRYAYCTVYCLGTILSMQISFVGFQPVLSHSHAAFCVFL
COIHAVDYLSKLNPOPEVLPFSVLSLVGVLLTVGALLMTGKISPMWGRFYSL
DPSANNNIPILASVSEHPTWSSYFPLGLLVFMPVGLVYCFNSLSDARFILMY
GVTSYFSAVMRLMLVLPVLCISGLVGSQVLSVSKNLDISRPKSKKQODSTY
PIKNEVASGMILVWAFITVYFTHSTWTSYSSPSIVLSARDGDGSRIFIDFRRA
YYLKHNTPEDAKVMWDYGYQITAMANTILVDNNTWNTNHSRQGVAMSTEBKA
YEMRELDVSVLVYFGLGTYSDDINKFLMWIRIGSDTDGKHENDAYTPTGEF
AVDRGSPVLNCLMYKCYRFGQVTEAKRPPGFDRVNAEIGNKDFELDLVEAY
TTEHVLVRIYKVDLNDRLSRT"

```

ORIGIN

Alignment Scores:
 Pred. No.: 5,12e-307 Length: 2293
 Score: 3647.00 Matches: 696
 Percent Similarity: 99.01% Conservative: 2
 Best Local Similarity: 98.72% Mismatches: 7
 Query Match: 98.65% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x HUMTMC (1-2293)

| | | | |
|----|-----|--|-----|
| QY | 1 | MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu | 20 |
| DB | 117 | ATGACTAAGTTGGATTTTGGATTTGGATTTGCTATGAGAGCAGACACTTTTGAAGCTT | 176 |
| QY | 21 | LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg | 40 |
| DB | 177 | CTCATCTGCAATGCTGCTGTATATCTTCTCCACTCGCTCTTTTGGTCTGCTGAGA | 236 |
| QY | 41 | PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu | 60 |
| DB | 237 | TTTGAAGGTTATCATCATGAGTTTGAATTCCTATTTTAAATATCGAGCTACCGATTCTG | 296 |
| QY | 61 | AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu | 80 |
| DB | 297 | TCTGAGGAGGGTTTATRAATTCACATACTGGTTGATGACCGAGCCTGTGACCTTTG | 356 |
| QY | 81 | GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr | 100 |
| DB | 357 | GSACGAATCATTCGAGGAAACAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC | 416 |
| QY | 101 | HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla | 120 |
| DB | 417 | CATGTACTCCATTTTTCACATCATCACATCGACATTCGGAATGCTGTGTGCTGCGC | 476 |
| QY | 121 | ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrIleGluLeuLysAsp | 140 |
| DB | 477 | CCTCTCTCTCCCTCCCTCACCATCGTCACATCCACTTACCAAGAGCTCAAGGAT | 536 |
| QY | 141 | AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg | 160 |
| DB | 537 | GCAGGGCTGGGCTTCTTGCTGCTGCAATGCTGTAGTCTTGGATATATCTCCGA | 596 |
| QY | 161 | SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysValLeuLeuThrTyr | 180 |
| DB | 597 | TCTGTGGCTGGCTCTATGATTAATGAAGGATTTGCCATCTTTTGGATGCTACTCACTAC | 656 |
| QY | 181 | TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu | 200 |
| DB | 657 | TACATGTGGATCAAGGAGTAAGACCTGGTTCCATCTGTGTGGCAGCTAAGTGGCTT | 716 |
| QY | 201 | AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuLeuLeuPro | 220 |
| DB | 717 | GCTTATTTTACATGCTGCTGCTATGCGGAGGTTATGTGTCTCTGATCACTTAATCTCT | 776 |
| QY | 221 | LeuHisValLeuLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr | 240 |
| DB | 777 | CTCCACGCTCTCGTGTGCTGCTCACAGGCGGTTTCTCTCCAGGATCTATGTGGCTAC | 836 |

| | | | |
|----|------|---|------|
| QY | 241 | CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln | 260 |
| DB | 837 | TGTACTGTTTACTGCTGGGCACATATCTTTCTATGCAGATCTCTCTTTGGTGTCCAG | 896 |
| QY | 261 | ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis | 280 |
| DB | 897 | CCTGTCTCTTTCATCAGACACATGGCAGCCTTTGGGTCTTTCGCAGATCCAT | 956 |
| QY | 281 | AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg | 300 |
| DB | 957 | GCTTTTGTGGATTACTGCGCAGCAGTGTGAATCCACACCAATTTGAAGTTCTTTCCGG | 1016 |
| QY | 301 | SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr | 320 |
| DB | 1017 | AGCGTCATCTCTGCTAGGCTTTGCTCTTCTCACCGTGGAGCTCTCTCATGCTGACA | 1076 |
| QY | 321 | GlyLysIleSerProTrpThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys | 340 |
| DB | 1077 | GGAAAAATATCTCCCTGGACGGGGCTTTCTACTGCTGCTGGATCCCTCTTATGCTAAG | 1136 |
| QY | 341 | AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr | 360 |
| DB | 1137 | ARCAACATCCCATCATCTGCTTCTGTGTGAGCATCAGCCACCAACCTGGTCTCTCATAC | 1196 |
| QY | 361 | TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer | 380 |
| DB | 1197 | TATTTTGACTCGAGCTCTCTGCTCATGTTTCCAGTTGGCTCTCTATCTGCTTTAGC | 1256 |
| QY | 381 | AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer | 400 |
| DB | 1257 | AACCTGCTGATGCCCGGATTTTATCATCATGATGATGTTGTTGTCACCGATGATGATTTCA | 1316 |
| QY | 401 | AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle | 420 |
| DB | 1317 | GCTGTAAATGGTGGCTTAATGCTAGTTGGACCTCTTATGTGCATTTCTCTGGCATT | 1376 |
| QY | 421 | GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys | 440 |
| DB | 1377 | GGAGTCTCCAGGTGCTGCTCCACATACATGAGAAATCTGACATAAGTCGCCACAGCAAG | 1436 |
| QY | 441 | LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet | 460 |
| DB | 1437 | AAGAGCAAGAACCAAGGATTTCCACTACCTTATTAAGAAATCAAGTGGCAAGTGGGATG | 1496 |
| QY | 461 | IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer | 480 |
| DB | 1497 | ATACTGTGTCATGGCTTTCTTTCTCATCACCTACACCTTTTCACTACCTGGGTGACCACT | 1556 |
| QY | 481 | GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyLysAspGlySerArgIle | 500 |
| DB | 1557 | GAGGCTTACTCTTCTCGTCCATTGCTATCTGCTGCTGATGGGATGGCAGTAGGATC | 1616 |
| QY | 501 | IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla | 520 |
| DB | 1617 | ATATTGATGACTTCGAGAGGATATATTGCTTCTGCTCATTAATACTCCAGNGGATGGC | 1676 |
| QY | 521 | LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle | 540 |
| DB | 1677 | AAGTCTATGCTGCTGGTGGATTTATGGCTATACAGATTACAGCTATGCAAAACCAAAAT | 1736 |
| QY | 541 | LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla | 560 |
| DB | 1737 | TTAGTGACAAATAACATGGAATAATACCATATTTTCTCGAGTAGGGCAGGCAATGGCG | 1796 |
| QY | 561 | SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal | 580 |
| DB | 1797 | TCCACAGAGGAAAGCCTATGAGATCATGAGGAGCTCGATGCTGAGTATGTGCTGGTC | 1856 |
| QY | 581 | IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal | 600 |
| DB | 1857 | ATTTTGGAGGCTCTACTGGGTATTTCTCTGATGATATCAACAGTTTCTTTGATGGTC | 1916 |

```

601 A-gilleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrPro 620
1917 CGGATGGAGGAGGACAGATACAGCAACATATCAAGGAGATGACTATATCTCCA 1976
621 ThrGlyGluPheArgValaspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
1977 ACTGGGGAGTCCGTTGGACCGTGAAGGTTCTCCAGTGTCTCAACTGCCTCATGTAC 2036
641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
2037 AGATGTGTACTATCGCTTTGGACAGGTTTACACAGAAAGCAAGCGCTCCAGGCTTT 2096
661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
2097 GACCGTGTCCGAATGCTGAGATTGGGAATTAAGACTTTGAGCTTGCTCGGAGAA 2156
681 GlyTyrThrThrGluHisThrLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
2157 GCATATACACAGAAATGCTGCTGTCAGGATATACAAAGTAAAGGACCTGGATAATCGA 2216
701 GlyLeuSerArgThr 705
2217 GGTGTGTCAAGGACA 2231

RESULT 10
HSM809254      2898 bp      mRNA      linear      PRI 30-AUG-2003
LOCUS          Homo sapiens mRNA; cDNA DKFZp79N2434 (from clone DKFZp79N2434).
DEFINITION     BX649102
ACCESSION      BX649102.1 GI:34368274
VERSION        1
KEYWORDS       Homo sapiens (human)
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2898)
Lauber, J., Bahr, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
Fobo, G., Han, M. and Wiemann, S.
The German Human cDNA Consortium
Direct Submission
Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp79N2434) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
source
1..2898
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp79N2434"
/tissue_type="liver"
/clone_lib="779 (synonym: hnccl). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="fetal"
polyA_signal
2817..2822
polyA_site
2863
ORIGIN
Alignment Scores:
Pred. No.: 1,29e-306 Length: 2898
Score: 3644.00 Matches: 695
Percent Similarity: 98.87% Conservatives: 2
Best Local Similarity: 98.87% Mismatches: 8
Query Match: 98.57% Indels: 0
DB: 9 Gaps: 0

```

```

US-10-028-384-12 (1-705) x HSM809254 (1-2898)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
186 ATGACTAAGTTTGGATTTTTCGATTTGCTCTATGAGAGAGGACACACTTTTGAAGCTT 245
21 LeuIleLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
246 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTCGTCTGTTGCTGCTGAGACA 305
41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
306 TTTGAAAGTGTATCCATGAGTTTGATCCGCTACTTTTAAATTATCGGACTACCAAGGTTCC 365
61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeu 80
366 GCTGAGGAGGGGTTTATAAATCCATAACTGGTTTGTATGATGACCGAGCTGTACCTTTG 425
81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
426 GACGGAATCATTTGGAGAAACAATTTACCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 485
101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
486 CATGTACTCAATTTTTCACATCACCATCGACATTCGGAATGCTGTGTGTTCTTGCCG 545
121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
546 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 605
141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
606 CGAGGGGCTGGGCTTCTTCTGCTGCCATGATGCTGTAGTTCTTGGATATATCTCCGA 665
161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
666 TCTGTGGTGGCTCTCTATGATAATGAGGGATTGTCATCTTTTGCATGCTACTCACTAC 725
181 TyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
726 TACATGTGGATCAGGCGAGTAAGACTGTTCCATCTGTTGGGAGCTAAGTGTGCCCTT 785
201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
786 GTTATTTCTCATGCTGCTCTCATGGGAGGTTATGTTTCTTCTGATCACTTAATCTCT 845
221 LeuHisValLeuValLeuMetLeuThrClyArgPheSerHisArgIleTyrValAlaTyr 240
846 CTCCACGTCCTCTGCTGCTGATGCTCACAGGCGTTTCTCTCACCGGATCTATGTGGCTAC 905
241 CystThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
906 TGTACTGTTTACTGCTGGGCACTATACTTTCTATGCAGATCTCTTTGTGGGTTTCAG 965
261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
966 CTTGTCTCTTCTCATCAGAGCACATGGGAGCTTTTGGGCTCTTTTGTCTCTGCGAGATCAT 1025
281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
1026 GCCTTTGTGATTACCTGGCAGCAAGTTGATCCACAACTTTTGAAGTTCTTTTCCGG 1085
301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
1086 AGCGTCATCTCTCTGTAGGCTTTGTCTCTCACCGTGGAGGCTCTCTCATGCTGCTGACA 1145
321 GlyLysIleSerProTyrThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
1146 GGAATAATCTCTCTGACGGGGCTCTCTACTCGCTGTGATCCCTCTTATGTGAAG 1205
341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrPseSerTyr 360
1206 AACCAACATCCCATCATTTGTTCTGTCTGAGCATCAGCCCAACACTGCTGCTCTCATAC 1265

```

QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1266 TATTTGACCTGACAGCTCCTCGCTTCATGTTTCCAGTTGGCCCTCTATTACTGCTTTAGC 1325
 QY 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1326 AACCTGCTGATGCCCGGATTTTATCATCATGATATGGTGTACCGACGATGACTTTTCA 1385
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1386 GCTGTAATGGTGGCTCTAATGCTAGTGTGGCACCTGTTATGTGCATTCTCTGGCATT 1445
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProAspIys 440
 Db 1446 GGAGTCTCCAGAGGCTGCTCCACATCATGAAGAATCTGGACATAGTCTGCACAGCAG 1505
 QY 441 LysSerLysLysGlnArgAspSerThrTyrProIleLysIleLeuValAlaSerGlyMet 460
 Db 1506 AAGAGCAAGAACACAGCATTCACCTACCTATTAAAGAAAGAGTGGCAAGTGGATA 1565
 QY 461 IleLeuValMetAlaPheLeuIleThrTyrPheHisSerThrTrpValThrSer 480
 Db 1566 ATACTGGTCATGGCTTCTTCTTCATCACCCTACCTTCATTCACCTGGGTGACCACT 1625
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1626 GAGGCTACTCCTCTCCGTCATGTACTATCTGCCGCTGGTGGGATGCCAGTAGGATC 1685
 QY 501 IlePheAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGluAspAla 520
 Db 1686 ATATTGTAGTCTCCAGAGCAATATTATGCTTCGTCATTAATCTCCAGAGATGG 1745
 QY 521 LysValMetSerTyrPyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1746 AAGGTCACTGCTCGTGGGATTAAGCTATACAGTATACAGTATGCAAAACCAAT 1805
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaValAla 560
 Db 1806 TTAGTGACATAACATGGAATATATCCCATTTCTCGAGTAGGCGGCAATGGCG 1865
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1866 TCCACAGAGGAAAAAGCCTATGATCATGAGGAGCTTGATGTACAGTATGTGCTGTC 1925
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 Db 1926 ATTTTGGAGCCCTCACTGGGTATTCCTCTGATGATATCAACAGTTTCTTTGGATGGTC 1985
 QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1986 CGGATGGAGGAGCACAGATACAGGCAACATATCAAGGAGAAATGACTATTATATCTCA 2045
 QY 621 ThrGlyGluPheArgValAspArgGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 2046 ACTGGGAGTTCCGTGTGGACCGTGAAGGTTCCTCAGTGTCTCAACTGCCTCATGATC 2105
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2106 AAGATGTGTACTATCGCTTGGACAGTTTACACAGAGCAAGCGTCTCCAGGCTTT 2165
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheCysLeuAspValLeuGluGlu 680
 Db 2166 GACCGTGTCCGAATGCTGAGATGGGAATAAAGACTTTGAGCTTGATGCTCTGGAGAA 2225
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2226 GCATATACCAAGAACATTTGGTGTGCTCAGGATATACAGGTAAGGACCTCGGATAATCA 2285
 QY 701 GlyLeuSerArgThr 705
 Db 2286 GCCTGTGTACAGGACA 2300

RESULT 11

CQ834710
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE

AUTHORS
 TITLE
 JOURNAL

FEATURES

source
 1..2435
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 73..2187
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05570.1"
 /translation="MTKFLRLSVEKQDTLLKLLLSMAAVLSFSLFAVLRFESV
 IHEFPYNYRTTFLAESEGYKPHNFDRAWYPLGRIIGGTYTTPGLMITSAAIYHV
 LHPFTIDIRNVCFVFLAPLSPSTTTIYVHLTKELDAGALLAAMIAVVPQYR
 SVAGSDNEGALFCMLTYVMKAVATGSGICAAKALAFYVWSGSGVVFILNI
 PLHVLMLTGRFPHRIYVAYCTVYCLGTLISMLQISFVGPVLSSHMAARGVGLC
 QHAFVYHRSKLNQFBEVLRVSVISUGFVLTVGALLMLTKISFWTGRYISLD
 PSYANNPIIASVSEHPTWSSYFDLQLLVFMFVGLYFCNSLSDARIFILMTG
 VTSMTFSVAMRLMLVAPVNCILSGISGVQLVTKNLDISRPDKSKKQDSTYP
 IKNEVASGMLVAFLLTYTFTSTWTSSEYSPSIVLSARGGDSRIIFDFPREAY
 YLNRNTPEDAKVMSWDYGYITAMANRTILVDNNTWNTTHSRVQQAASSTEEKAY
 EIMRELDVSYLVIFEGTLGYSDDDINKFLMWRIIGSTDTGTHIKENDVYTPGEPR
 VDREGSVLLNCLMYKCYRFGQVYTEAKRPPGDFRVARNAEIGNKDFELDLBEAYT
 TEHLVRIYKVKLDLNRGLSRT"

CDS

1..2435
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 73..2187
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05570.1"
 /translation="MTKFLRLSVEKQDTLLKLLLSMAAVLSFSLFAVLRFESV
 IHEFPYNYRTTFLAESEGYKPHNFDRAWYPLGRIIGGTYTTPGLMITSAAIYHV
 LHPFTIDIRNVCFVFLAPLSPSTTTIYVHLTKELDAGALLAAMIAVVPQYR
 SVAGSDNEGALFCMLTYVMKAVATGSGICAAKALAFYVWSGSGVVFILNI
 PLHVLMLTGRFPHRIYVAYCTVYCLGTLISMLQISFVGPVLSSHMAARGVGLC
 QHAFVYHRSKLNQFBEVLRVSVISUGFVLTVGALLMLTKISFWTGRYISLD
 PSYANNPIIASVSEHPTWSSYFDLQLLVFMFVGLYFCNSLSDARIFILMTG
 VTSMTFSVAMRLMLVAPVNCILSGISGVQLVTKNLDISRPDKSKKQDSTYP
 IKNEVASGMLVAFLLTYTFTSTWTSSEYSPSIVLSARGGDSRIIFDFPREAY
 YLNRNTPEDAKVMSWDYGYITAMANRTILVDNNTWNTTHSRVQQAASSTEEKAY
 EIMRELDVSYLVIFEGTLGYSDDDINKFLMWRIIGSTDTGTHIKENDVYTPGEPR
 VDREGSVLLNCLMYKCYRFGQVYTEAKRPPGDFRVARNAEIGNKDFELDLBEAYT
 TEHLVRIYKVKLDLNRGLSRT"

ORIGIN

Alignment Scores:
 Pred. No.: 8,29e-306 Length: 2435
 Score: 3633.50 Matches: 696
 Percent Similarity: 98.87% Conservative: 1
 Best Local Similarity: 98.72% Mismatches: 7
 Query Match: 98.28% Indels: 1
 DB: Gaps: 1

US-10-028-384-12 (1-705) x CQ834710 (1-2435)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 73 ATGACTAAGTTTGGATTTTGGATTTGCTTATGAGAGCAGGACACACTTTTGAAGCTT 132
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 133 CTCATTCTGTCAATGGCTGCTGCTATTTATTCCTTCTCCACTGCTGTTGCTGCTGAGA 192
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 193 TTTGAAGTGTATCCATGAGTTTGCATCGTACTTTATTTATTCGAGCTACCCAGGTCTCTG 252
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
 Db 253 GCTGAGAGGGGGTTTATTAATTTCCATACTGGTTTGTATGACCGAGCTGTGACCTTTG 312
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 313 GGACGAATCATTCGGAGGAACAATTTACCCAGGTTTATATGATCACCTCTGCTGCAATCTAC 372
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120

CQ834710 2435 bp DNA linear PAT 29-JUL-2004
 Sequence 581 from Patent WO2004058805.
 CQ834710
 CQ834710.1 GI:50834247
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. Matsuda, A. and Yoneta, S.
 T cell activating Gene
 Patent: WO/2004058805-A 581 15-JUL-2004;
 Asahi Kasei Pharma Corporation (JP)
 Location/Qualifiers
 1..2435

/organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 73..2187
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05570.1"
 /translation="MTKFLRLSVEKQDTLLKLLLSMAAVLSFSLFAVLRFESV
 IHEFPYNYRTTFLAESEGYKPHNFDRAWYPLGRIIGGTYTTPGLMITSAAIYHV
 LHPFTIDIRNVCFVFLAPLSPSTTTIYVHLTKELDAGALLAAMIAVVPQYR
 SVAGSDNEGALFCMLTYVMKAVATGSGICAAKALAFYVWSGSGVVFILNI
 PLHVLMLTGRFPHRIYVAYCTVYCLGTLISMLQISFVGPVLSSHMAARGVGLC
 QHAFVYHRSKLNQFBEVLRVSVISUGFVLTVGALLMLTKISFWTGRYISLD
 PSYANNPIIASVSEHPTWSSYFDLQLLVFMFVGLYFCNSLSDARIFILMTG
 VTSMTFSVAMRLMLVAPVNCILSGISGVQLVTKNLDISRPDKSKKQDSTYP
 IKNEVASGMLVAFLLTYTFTSTWTSSEYSPSIVLSARGGDSRIIFDFPREAY
 YLNRNTPEDAKVMSWDYGYITAMANRTILVDNNTWNTTHSRVQQAASSTEEKAY
 EIMRELDVSYLVIFEGTLGYSDDDINKFLMWRIIGSTDTGTHIKENDVYTPGEPR
 VDREGSVLLNCLMYKCYRFGQVYTEAKRPPGDFRVARNAEIGNKDFELDLBEAYT
 TEHLVRIYKVKLDLNRGLSRT"

Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzinski, M.I., Skalek, U., Smalls, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903. (2002)

12477932

2 (bases 1 to 2736)

Strausberg, R.

Direct Submission

Submitted (13-SEP-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaps@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)

Gaithersburg, Maryland

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Akhtar, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Higgins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAC Plate: 81 Row: n Column: 21

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6680497.

Location/Qualifiers

1. .2736

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="MGC:46919 IMAGE:5353797"

/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."

/clone_lib="NCI CGAP_Mam6"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

1. .2736

/gene="Itml"

/db_xref="LocusID:16430"

/db_xref="MG:105124"

108. .2225

/gene="Itml"

/codon_start=1

/product="integral membrane protein 1"

/protein_id="AAH37612.1"

/db_xref="GI:22902289"

/db_xref="LocusID:16430"

/db_xref="MG:105124"

/translation="MTKLGLRLSYEKQDTLLKLLISMAVLFSFRLFAVRFESV IHEPDPVNTVTRFLAEGFYFHNWDDRAWPLGRIGTIVYPLGLMITAAIYHV LHPHIDIRNVCFIAPLPSFTTIVTHTKLKADAGALLAAIAVVPVYSR SVAGSNEGIAIFCMLLTYMKIKAVTGSIVYAAKALAFVYSSKGVYFLINL IPEHLVLMITGRFSRIYVAYCTVYCLGILISNQISFVGFPQPLSSEHMAAFVGL CQIHAFVDYLRSLNPOQFEVLFRFSLVSLVGLVLTIVGALLMUTGKISFWTRFVSL

ORIGIN

Alignment Scores:

Pred. No.: 1,08e-305 Length: 2736

Score: 3633.00 Matches: 694

Percent Similarity: 98.87% Conservative: 3

Best Local Similarity: 98.44% Mismatches: 8

Query Match: 98.27% Indels: 0

DB: 10 Gaps: 0

US-10-028-384-12 (1-705) x BC037612 (1-2736)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20

Db 108 ATGACTAAGCTTGGATTTTGGCATTTGCTCTATGAGAGCAGACACACTTCTAAGCTT 167

Qy 21 LeuLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40

Db 168 CTATCTCTGTCGATGGCTGCTGTTGTTATCTTTTCTACTCTCTTTTCTGCTGCTGAGA 227

Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60

Db 228 TTTGGAAGTGTATCATCCATGATGTTGATCGGTACTTTTAAATTATCGGACTACCGGTTCTG 287

Qy 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTrpTyrProLeu 80

Db 288 GCTGAGAGGGGTTTATAAATCCATACTGTTGATGACCGGCTTGTACCCCTTTC 347

Qy 81 GlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100

Db 348 GGCAGATCATTTGGAGGACAAATTACCCAGGTTTAAATGATCACTTCTGCTGCAATCTAC 407

Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120

Db 408 CATGTACTCCATTTCTTCCATATCATTAATGACATTCGGAATGCTGTGTTTCTCGGCC 467

Qy 121 ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140

Db 468 CCATTTTCTTCCTTTCACCACTGTTACGCTACCACTTACCAAGAGCTCAAGGAT 527

Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160

Db 528 GCAGAGCTGGCTTCTTGTGCTGCCATGATGCTGTATGCTTCTGGGTATATTTCTCGA 587

Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180

Db 588 TCTGTAGTGTCTCTATGATATGAAGAAATGCTATCTTTTCATGCTGCTTACTTAC 647

Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaLeu 200

Db 648 TACATGTGGATCAAGGAGCTGAGACTGGTTTCCATCTATTTGGGCTGCAAGTGTGCCCTC 707

Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220

Db 708 GCTTATTCTACATGGTCTCTTCAATGGGAGGCTATGTTGTTCTGATCAACTGATCTCT 767

Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240

Db 768 CTACATGTCCTGGTGTCTAATGCTGACAGCGCGTTTTTCTCACCGGATCTACGTAGCTAC 827

Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260

Db 828 TGTACTGTCTACTCCCTGGGACCATTCTTCTATGACAGATTTCTTGTGTTGTTTCCAG 887

Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280

888 CCGTCCTTTTCATCAGAACACATGGCCTTTGGAGTGTGTTGGTCTCTGTCAGATCCAT 947
281 AlaPheValAspTyrLeuArgSerIysLeuAsnProGlnGlnPheGluValLeuPheArg 300
948 GCTTTTCGTAGATTACCTCGCAGCAAGTGAATCAGACAGCAATTCGAAGTCTCTTTCCGG 1007
301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
1008 AGTGTAITCTCCCTGGTGGCTTTGCTCTCTCTCACTGTGGAGCTCTCTCTATGCTAACA 1067
321 GlyIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
1068 GGAATAATTTCTCCCTCGACAGGGCGTTTCTACTCTCTGCTGGATCCCTCTTATGCTAAG 1127
341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTyrSerTyr 360
1128 AATRAACATTTCCATATTGTCATCTGTTCTGAGCAGCCAGCCACACTGGTCTTCTCTAC 1187
361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
1188 TATTTTGATCTACAGCTCTTGTCTCTCATGTTTCAGTTGGCTCTATTACTGCTTTAGC 1247
381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
1248 AACCTGCTGATGCTCGGATTTTATCATCATGATGATGATGATGATGATGATGATGATGAT 1307
401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
1308 GCTGTAATGGTGGCTCTAATCTGGTATGTCACCTGTTATGTCATTCCTTCTGGCAT 1367
421 GlyValSerGlnValLeuSerThrTyrMetIleSerLeuAsnIleSerArgProAspLys 440
1368 GGTGTTTCCAGGTCGTGTCATATATGAAATCTGACATAGTCCGCCAGACAAG 1427
441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
1428 AAGAGCAAGAAGCAACAGGATCTTACCTTATTAAGATGAGGTGGCGATGGGATG 1487
461 IleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTyrValThrSer 480
1488 ATACTGTCATGGCTTTTCTCATCACCTACACGTTTCACTTCGATTCGATTCGATTCGATTC 1547
481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
1548 GAAGCCTATTCTTCTCCCTCCATGTAATGCTGCTGCTGGTGGGATGGCAGTAGATC 1607
501 IlePheAspAspPheArgGluAlaTyrTyrThrLeuArgHisAsnThrProGluAspAla 520
1608 ATTTTTCATGACTTCCGAGAGCGTATATTGGCTCGTCCGTCACCAATCTCCAGAGGATGCA 1667
521 LysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaSerThrIle 540
1668 AAGTCATGTCATGTTGGATTTATGGCTACCAATTTACTGCAATGGCAATCGCAAT 1727
541 LeuValAspAsnAsnThrThrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
1728 TTAGTGACAATAACACATGGAATAATACCAATATTTCTGAGTAGCGAGCAATGGCA 1787
561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
1788 TCCACAGAGAAAGCCTATGAATCATGAGGAGCTTGATGTCAGCTATGCTGCTGTC 1847
581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAlaAsnLysPheLeuTyrMetVal 600
1848 ATTTTGGAGGCTTACTGGTATTCTTCGATGATATCAACAGTTTCTTTGGATGTC 1907
601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
1908 CGATTGGAGAGACAGACAGAGAGACACATTAAGAGAAATGACTACTATATCTCT 1967
621 ThrGlyCysPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
1968 ACTGGGAATTCGTGTTGATCGTGAGGGTCTCCCGGCTGCTGCTCACTGCTTATGATC 2027

641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
2028 AAAATGTGTACTACCGCTTTGGCAGGCTCTACACAGAAGCAAGCGTCCACAGGCTTT 2087
661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
2088 GACCGTGTTCGAATGCTGAGATTGGTAATAAAGACTTTGAGCTTGATGCTCTGGAGAA 2147
681 GlyTyrThrThrGluHisTyrLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
2148 GCGTATACACAGAACACTGGCTAGTCAGATATACAGGTAAAGGACCTGGATATATCGA 2207
701 GlyLeuSerArgThr 705
2208 GGCTTGTCAAGGACA 2222
RESULT 13
AX799090
LOCUS AX799090 3094 bp mRNA linear PAT 10-OCT-2003
DEFINITION Sequence 9 from Patent WO03054008.
ACCESSION AX799090
VERSION AX799090.1 GI:37605062
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 3094
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/note="ITM1 accession No NM-008408"
ORIGIN
Alignment Scores:
Pred. No.: 1.27e-305 Length: 3094
Score: 3633.00 Matches: 694
Percent Similarity: 98.87% Conservative: 3
Best Local Similarity: 98.44% Mismatches: 8
Query Match: 98.27% Indels: 0
DB: 6 Gaps: 0
US-10-028-384-12 (1-705) x AX799090 (1-3094)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 112 ATGACTAGCTGGATTTTTCGATTGCTCTATGAGAGAGAGACACACTTCTTAAAGCTT 171
QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 172 CTCATCTCTGTCGATGGCTGCTGTGTATCTTTTCTACTCTGCTTTTGTGCTGCTGAGA 231
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 232 TTTGAAAGTGTCATCCATGAGTTTGAATCCGCTTATTAATTCGAGCTACCCGGTTTCTG 291
QY 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeu 80
DB 292 GCTGAGAGGGGCTTTTATAATTCATACTGTTTATGATGATGATGATGATGATGATGATGAT 351
QY 81 GlyArgIleLeuGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 352 GCGCAATCATTTGAGGAGAAATTTACCCAGGTTTATGATCACTTCTGCTGCAATCTAC 411
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120

Db 412 CATGTACTCCATTTCTTCATATCACTATTGACATTCGAAATGCTGTGTTTCTCGCC 471
 Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 472 CCACATTTTCTCTCTTTTCCACCAATCGTTAGTACACACCTTACCAAGAGCTCAAGAT 531
 Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
 Db 532 CGAGAGCTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591
 Qy 161 SerValAlaGlySerTyrAspGlnGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 592 TCTGTAGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651
 Qy 181 TyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
 Db 652 TACATGTGATCAAGCAGTGAAGCTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 711
 Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuPro 220
 Db 712 GCTTATTTCTACATGCTCTTCTCATGGGAGCTATGTTCTTCTGATCACTTGATTTCT 771
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 Db 772 CTACATGCTGCTGTGTAATGCTGACAGCCGCTTTTCTCACCGGATCTAGTAGCCCTAC 831
 Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 832 TGTACTGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
 Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLysCysGlnIleHis 280
 Db 892 CCGTCTCTTTCATCAGAACACATGCGAGCTTTGGAGTGTGGTCTCTGTCAGATCCAT 951
 Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 952 GCITTCGTAGATACCTGGCAGCAAGTGAATCCACAGCAATTCGAAGTCTTTTCCGG 1011
 Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyValAlaLeuLeuMetLeuThr 320
 Db 1012 AGTGTTATCTCCCTGGTGTGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1071
 Qy 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 1072 GGAAATTTCTCCCTGGACAGGCGTTTCTACTCTCTCTCTCTCTCTCTCTCTCTCT 1131
 Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrSerTyr 360
 Db 1132 AATAACATTCCTATTTATGATCTGTTTCTGACACAGCCACACCTGGTCTTCTCTAC 1191
 Qy 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1192 TATTTGATCTACAGCTCTTCTTCTCATGTTTCCAGTGGCTCTATTACTGCTTTAGC 1251
 Qy 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1252 AACCTGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311
 Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1312 GCTGTATGTCGCTCTAAAGTGTGATTTGGACCTGTTATGTCATTTCTTCTGGCAT 1371
 Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1372 GGTGTTTCCAGAGTGTGCTGCACATATGAAATCTGGACATAAGTCCGCCAGACAG 1431
 Qy 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1432 AAGAGCAAGACAGACAGATTCCTTACCTTATTAAGATGAGTGGGAGTGGGATG 1491
 Qy 461 IleLeuValMetAlaPheLeuIleThrTyrThrPheHisSerThrTyrValThrSer 480

Db 1492 ATACTGGTCATGGCTTTTCTTCATCACCTACACGTTTCTTCTGACTTGGGTGACCACT 1551
 Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 Db 1552 GAAGCCTATTTCTTCCCTCCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1611
 Qy 501 IlePheAspPheArgGluAlaTyrTyrTyrTyrLeuArgHisAsnThrProGluAspAla 520
 Db 1612 ATTTTGTATGATCTTCGAGAGCGTATTATTGGTCCCTCACAATCTCTCCAGAGATGCA 1671
 Qy 521 LysValMetSerTyrTyrPheTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1672 AAGTCTCATGCTGCTGGGATTTATGGCTACCAAAATTTACTGCAATGGCAATTCGCA 1731
 Qy 541 LeuValAspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1732 TTAGTGGCAATTAACATGGAATTAATACCAATATTTCTCGAGTAGGCGAGCAATGCA 1791
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1792 TCCACAGAAAGAAAAGCCTATCAAAATCATGAGGAGCTTGTATGCTGCTGCTGCT 1851
 Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetVal 600
 Db 1852 ATTTTGGAGGCTTACTGGGTATTTCTCGATGATATCAACAGTTTCTTGGATGGTC 1911
 Qy 601 ArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1912 CGGATTTGGAGGAGCAGACAGACAGACACATTAAGGAGATGACTACTATCTCTCT 1971
 Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1972 ACTGGGCAATTCGCTGTTGATCGTGGGTTTCTCGGCTCTCTCACTGCTTTATGATC 2031
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2032 AAAATGTGTTACTACCGCTTTGGCAGGCTCTACACAGAACCAAGCGTCCACAGGCTT 2091
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2092 GACCGTTCGAAATGCTGAGATTGTAATAAGACTTTGAGCTTGATGCTCTGAGGAA 2151
 Qy 681 GlyTyrThrThrGluHisTyrPheValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2152 GCGTATACCAAGAACACTGGCTAGTCAGGATATACAGGTAAAGACCTGGATATCGA 2211
 Qy 701 GlyLeuSerArgThr 705
 Db 2212 GGCTTTGTCAAGGACA 2226

RESULT 14
 MUSTRPR
 LOCUS Mus musculus integral membrane protein 1 (Itim1) mRNA, complete cds.
 DEFINITION
 L34260
 ACCESSION
 L34260.1 GI:508542
 VERSION
 transmembrane protein.
 KEYWORDS
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 3094)
 AUTHORS
 Hong, G., Deleerijder, W., Kozak, C.A., Van Marck, E., Tylzanowski, P.
 and Mergaert, J.
 TITLE
 Molecular cloning of a highly conserved mouse and human integral
 membrane protein (Itim1) and genetic mapping to mouse chromosome 9
 Genomics 31 (3), 295-300 (1996)
 JOURNAL
 MEDLINE
 PUBMED
 96435424
 FEATURES
 Location/Qualifiers
 1..3094
 /organism="Mus musculus"
 /mol_type="mRNA"

```

/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/chromosome="9"
/tissue_type="mandibular condyle"
1. .3094
/gene="Itm1"
1. .111
/gene="Itm1"
112. .2229
/gene="Itm1"
/notes="putative"
/codon_start=1
/product="integral membrane protein 1"
/protein_id="A047775.1"
/db_xref="GI:508543"
/translation="MYKLGFLRLSYEKQDTLLKLLLSMAAVLSFSTRFLPAVLRPESV
LHEFDIFNRTFLAEGFYKFRHWFDDRAWYPIGRIIGGTIYPLGMITSAAIYHV
LHFPHTIDIRNVCVFLAPLFSSFTTIVTHTLKDKAGLLAAAMIAVPGYISR
SVAGSDNEGIAIFCMLLTYMMKAVKTGSIYMAAKCALAFYVMSWGCVFLINL
IPHVAVIDLGRFSHRIYVAYCTVYCLGTILSMQISFVGFPVLSSHMAAFVFLG
COLHAFVDLRLSKLNPOQPELVPSVSLVGLLVGALLMLTGKISPTWGRFYSL
DPSYAKNNIPILIASVEHOPTWSSVYFDLQLLVFEPVGLYVCSNLSDAFIFIMY
GVTSMVESAWRLMLVLAFCILSGISGVLSVYMKNLDIRPDKSKKQDQSTY
PIKNEVASGMLVMAEFLITYTPHSTWVTSYSEYSPSIVLSARGGDSRIIFDDPREA
YWLHNTPEDAKMSWDYQGITAMANTILVDNNTWNTHISRVGQAMASTBEKA
YEIMRLDYSYLVIFVGLTGYSDDINKFLMWYRIGGSTETGRHKENDYYTPTGEF
VRDEGSPVLLNCLMYKVCYRFGQVYTEAKRPPGDFRVNABIGNKDFELDLSEAY
TTEHMLVRIYKVKDLNRLSRT"
112. 201
/gene="Itm1"
202. .2226
/gene="Itm1"
/product="transmembrane protein"
/notes="putative"
2230. .3094
/gene="Itm1"
2443. .2448
/gene="Itm1"
/notes="putative"
2470
/gene="Itm1"
/notes="putative"
2704. .2709
/gene="Itm1"
/notes="putative"
2723. .2728
/gene="Itm1"
/notes="putative"
2725
/gene="Itm1"
/notes="putative"
2738
/gene="Itm1"
/notes="putative"
3075. .3080
/gene="Itm1"
/notes="putative"
3093
/gene="Itm1"
/notes="putative"

sig_peptide
mat_peptide
3'UTR
polyA_signal
polyA_site
polyA_signal
polyA_signal
polyA_site
polyA_site
polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 1-27e-305 Length: 3094
Score: 3633.00 Matches: 694
Percent Similarity: 98.87% Conservative: 3
Best Local Similarity: 98.44% Mismatches: 8
Query Match: 98.27% Indels: 0
DB: 10 Gaps: 0

US-10-028-384-12 (1-705) x MUSTRRP (1-3094)

```

1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
112 ATGACTAAGCTTGGATTTTTCGATTGCTTATGAGAGCAGCAGACACTTCTAAAGCTT 171
21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
172 CTCATCTCTGTCGATGGCTGCTGTGTATCTTTTCTACTCTCTCTTTTGTGCTGAGA 231
41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
232 TTTGAAAGTGTCAATCATCGATGTTGATCCGACTTAAATATATCGGACTACCCGGTTCTG 291
61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaIleTyrProLeu 80
292 CTTGAGGAGGGGTTTATAAATTCATAAATCGGTGTGATGACCGGGCTTGGTACCTTTG 351
81 GlyArgIleLeuGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
352 GGCAGATCATTTGGAGGAACAATTTACCCAGGTTAATATGATCACTTCTGCTGCAATCTAC 411
101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
412 CATGTACTTCCATTTCTTCCATATCACTATTGACATTTCGAAATGCTGTGTTTTCTCGCC 471
121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
472 CCACATTTCTCTCTTTTCAACCACCATCGTTACGTACCACTTTACCAAAGAGCTCAAGGAT 531
141 AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
532 GCAGAGCTGGGCTTCTTGTCTGCTGCTGATGATGCTGTAGTCTCTGGGTATATTCTCGA 591
161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
592 TCTGTAGCTGCTCTCTATGATAATGAAGGATTCCTATCTTTGCAATGCTGCTTACTTAC 651
181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
652 TACATGTGGATCAAGGAGCTGGAAGACTGGTTCCTATTTGGGGCTGCAAGTGTGCGCTC 711
201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
712 GCTTATTCTACATGCTCTCTTCATGGGAGGCTATGTGTCTCGATCAACTGATTCTCT 771
221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
772 CTACATGCTCTGTGTCTAATGCTGACAGGCGGTTTTTCTCACCGGATCTACGTAGCCTAC 831
241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
832 TGTACTGTCTTACTGCTGGGACCATCTTCTATGACAGATTTCTTTTGTGTTTCCAG 891
261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
892 CCCGCTCTTTCATCAGAACACATGGCAGCCTTTGGAGTGTGTGCTCTCTGATGATCCAT 951
281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
952 GCTTCTGTAGATTACCTGCGCAGCAAGTTGAATCCACAGCAATTCGAAGTCTCTTTCCGG 1011
301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
1012 AGTGTATCTCCCTGGTGGCTTGTCTCTCTCCTCCTGCTGGAGCTCTCTCTCATGCTAACA 1071
321 GlyLysIleSerProTrpThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
1072 GGAATAATTTCTCCCTGGACAGGCGGTTTCTACTCTCTGCTGGATCCCTCTTATGCTAAG 1131
341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
1132 AATAACATTTCCATATTGATCTGTTTCTGAGACCCAGCCCAACCTGCTCTCTCTAC 1191
361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380

LOCUS CQ492624 2760 bp DNA linear PAT 30-JAN-2004
DEFINITION Sequence 24491 from Patent WO0160860.
ACCESSION CQ492624
VERSION CQ492624.1 GI:41458243
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.: 1,628-305 Length: 2760
Score: 3631.00 Matches: 698
Percent Similarity: 98.87% Conservative: 1
Best Local Similarity: 98.73% Mismatches: 6
Query Match: 98.21% Indels: 2
DB: 6 Gaps: 0
US-10-028-384-12 (1-705) x CQ492624 (1-2760)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 105 ATGACTAAGTTGGATTGTTTGGATTGCTTATGAGAGCAGGACACACTTTTGAAGCTT 164
QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr 40
DB 165 CTCATTCTGCTCAATGGCTGCTGTTATCTCTCCATCGTCTGTTTGTCTGCTGAG 224
QY 40 gPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLe 60
DB 225 ATTGGAAGTGTTATCCATGAGTTTGAATCGTACCTTTAATTATCGACTACCAAGTTCT 284
QY 60 uAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTyrPyrProLe 80
DB 285 GGCTGAGGAGGGGTTTATAAATTCATAAATTCGATGAGTGGTTCATGACCGAGCTGGTACCCCTT 344
QY 80 uGlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 345 GGAGCAATCATTTGAGGAGCAATTTACCCAGGTTTATGATCAGCTCTGCTGCATCTA 404
QY 100 rHisValLeuHis-PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuA 120
DB 405 CCATGTACTCCATTTTTCATCCATCCATCGACATTCGGAATGCTGTGTGTCTCTGG 464
QY 120 lAproLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysA 140
DB 465 CCCCT 524
QY 140 sAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
DB 525 ATGACGGGCTGGGCT 584
QY 160 rSerValAlaGlySerTyrAspAsnGluGlyValIlePheCysMetLeuLeuThrT 180
DB 585 GATCTGTGGCTGGCT 644
QY 180 YrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaL 200
DB 645 ACTACATGTGGATCAAGGAGGAGTAAAGACTGGTTCATCTGTTGGGAGGAGTAAAGTGTGCC 704
QY 200 euAlaTyrPheTyrMetValSerTyrPyrGlyGlyTyrValPheLeuIleAsnLeuIleP 220

Db 705 ITGCTTATTTCTACATGCTCTCGTCAATGGGAGGTATGTTCCCTGATCACTTAATTC 764
Qy 220 roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaI 240
Db 765 CTCCTCCACGCTCTCGTCTGATGCTACAGGCGGTTTCTCTCACCGGATCTATGTGGCCT 824
Qy 240 YrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheG 260
Db 825 ACTGTACTGTTACTGCTGGGCACTATATCTTCTATGCAAGTCTCTTTTGGGTTTC 884
Qy 260 InProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleH 280
Db 885 AGCCTGTCTCTTTCATCAGACACATGCGAGCCTTTGGGGTCTTTGGTCTCTGCGACATCC 944
Qy 280 isAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheA 300
Db 945 ATGCCTTTGGATTACTCTGGCAGCAAGTTGAATCCACAACAATTTGAAGTTCTTTTTC 1004
Qy 300 rGserValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuT 320
Db 1005 GGAGCGTCACTCTCTGTAAGCTTTTGTCTCTCACCGTGGGAGCTCTCTCATGCTGA 1064
Qy 320 hrGlyLysIleSerProTTPThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaI 340
Db 1065 CAGAAATAATCTCCCTGACCGGGGCTTTCTACTCGTGTGGATCCCTCTTATGCTA 1124
Qy 340 ysAsnAnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerT 360
Db 1125 AGAACACATCCCATCTGCTCTGTCCTGAGCATCAGCCACACACTGGTCTCAT 1184
Qy 360 YrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheS 380
Db 1185 ACTATTGTGACCTGCACTCTCGTCTTCATGTTTCCAGTTGGCCTCTATTACTGCTTA 1244
Qy 380 erAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheS 400
Db 1245 GCACCTGCTCTGATGCCGATTTTATCATCATGATATGTTGACACCATGATCTTTT 1304
Qy 400 erAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyI 420
Db 1305 CAGCTGTAAATGGTGGCTCTAATGCTAGTGTGGCAGCTGTATGTCATTCCTCTGGCA 1364
Qy 420 leGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspI 440
Db 1365 TTGGAGTCTCCAGGTCGCTGCCACATACATGAAGAATCTGGACATAAGTCTGCCAGACA 1424
Qy 440 ysLysSerLysLysGlnInAspSerThrTyrProIleLysIleGluValAlaSerGlyW 460
Db 1425 AGAAGAGAGAAAGCAACAGGATTCACCTTACCTTATTAAGAATGAAGTGCAAGTGGA 1484
Qy 460 etIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrS 480
Db 1485 TGATACCTGGTCATGGCTTTCTTCTCATCACCTACACCTTTTCATCAACCTGGGTGACCA 1544
Qy 480 erGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgI 500
Db 1545 GTGAGGCTTACTCTTCTCGCTCCATTTGTTCTGCTGCTCCCGTGGTGGGATGCGAGTAGGA 1604
Qy 500 leIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspA 520
Db 1605 TCATATTGATGACTTCCGAGAGACATATTATTGGCTTCGTTCATAATCTCTCCAGAGGATG 1664
Qy 520 lalysValMetSerTrpTTPAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrI 540
Db 1665 CGAAGTCTATGCTCTGGTGGGATTAATGGCTATCAGATTACAGTATGGCAACCGAACAA 1724
Qy 540 leLeuValAspAsnAnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetA 560
Db 1725 TTTTAGTGACATAACACATGAATAATACCATATTTCTCGAGTAGGCGAGGCAATGG 1784
Qy 560 laSerThrGluGlyLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuV 580

Db 1785 CGTCCACAGAGGAAAAAGCCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTGCTGG 1844
Qy 580 alIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetV 600
Db 1845 TCAATTTTGGAGGCTCACTGGTATTCTCTGATGATATCAACAAGTTTCTTTGGATGG 1904
Qy 600 alArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrP 620
Db 1905 TCCGATTGGAGGGAGCACAGATACAGCAAAACATATCAAGAGAAATGACTATTATATCT 1964
Qy 620 roThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetT 640
Db 1965 CAACCTGGGAGTTCGCTGTGACCGTGAAGGTTCTCCAGTCTGCTCACTGCCTCATGT 2024
Qy 640 YrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaIleArgProProGlyP 660
Db 2025 ACAAGATGTGTACTATCTGCTTTGGACAGGTTTACACAGAAAGCCAGCGCTCCAGGCT 2084
Qy 660 heAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluG 680
Db 2085 TTGACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATCTCTGGAGG 2144
Qy 680 luGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnA 700
Db 2145 AAGCATATACACAGAACATTTGGCTGGTCAAGATATACAGGTAAGAGACCTGGATATC 2204
Qy 700 rGlyLeuSerArgThr 705
Db 2205 GAGGCTTGTCAAGGACA 2221

Search completed: December 15, 2004, 06:01:36
Job time : 7905.52 secs

Blank
Page

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:28:05; Search time 812.511 Seconds
(without alignments)
4554.822 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKGFFLSEKQDTLLKL.....HMLVRIYKVLDLNRGLSRT 705

Scoring table:

BLOSUM62
Gapop 10.0, Xgapext 0.5
Gapop 10.0, Ygapext 0.5
Gapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10028384/runat 14122004 131518.8007/app query.fasta_1.3740
-DB=N Geneseq 23Sep04 -QWTF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MAFIX=blotum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR VAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcg -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384.cgn 1.1.1986.grunat 14122004 131518.8007 -NCFU=6 -ICFU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N_Geneseq_23Sep04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002as: *
7: Geneseqn2002bs: *
8: Geneseqn2003as: *
9: Geneseqn2003bs: *
10: Geneseqn2003cs: *
11: Geneseqn2003ds: *
12: Geneseqn2004as: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|--------|-------|--------|-------------|--------------------|
| 1 | 3697 | 100.0 | 2472 | 10 ADD94793 | Add94793 Human ITM |
| 2 | 3697 | 100.0 | 2472 | 10 ADH28844 | Adh28844 Human chr |
| 3 | 3633 | 98.3 | 3094 | 10 ADD94791 | Add94791 Mouse ITM |
| 4 | 3631 | 98.2 | 2760 | 5 ABV24502 | Abv24502 Human pro |
| 5 | 2903.5 | 78.5 | 2855 | 4 ABL02795 | Ab102795 Drosophil |
| 6 | 2602.5 | 70.4 | 6153 | 4 ABL02794 | Ab102794 Drosophil |

| | | | | | |
|----|--------|------|------|-------------|--------------------|
| 7 | 2263.5 | 61.2 | 1371 | 12 ADE28508 | Adp28508 Human sec |
| 8 | 2247 | 60.8 | 2481 | 10 ADD94783 | Add94783 Human SIM |
| 9 | 2237 | 60.5 | 2710 | 10 ADD94785 | Add94785 Mouse SIM |
| 10 | 2177.5 | 58.9 | 2417 | 10 ADD94789 | Add94789 Drosophil |
| 11 | 2177.5 | 58.9 | 2699 | 4 ABL13247 | Ab113247 Drosophil |
| 12 | 2103 | 56.9 | 4922 | 4 ABL13246 | Ab113246 Drosophil |
| 13 | 2070.5 | 56.0 | 2256 | 6 ABT2036 | Abt2036 candida a |
| 14 | 2010 | 54.4 | 2232 | 6 ABT20789 | Abt20789 Aspergill |
| 15 | 1930 | 52.2 | 2466 | 10 ADB69900 | Adb69900 C. neofor |
| 16 | 1920 | 51.9 | 2733 | 10 ADD94787 | Add94787 Yeast STT |
| 17 | 1867.5 | 50.5 | 2000 | 10 ACC61293 | Acc61293 Gene sequ |
| 18 | 1867.5 | 50.5 | 2000 | 10 ADK63659 | Adk63659 Disease t |
| 19 | 1851 | 50.1 | 2603 | 8 ABT20191 | Abt20191 Aspergill |
| 20 | 1851 | 50.1 | 3969 | 8 ABT17781 | Abt17781 Aspergill |
| 21 | 1851 | 50.1 | 4603 | 8 ABT19595 | Abt19595 Aspergill |
| 22 | 1722 | 46.6 | 1848 | 8 ABT18969 | Abt18969 Aspergill |
| 23 | 1682 | 45.5 | 2882 | 10 ADB69539 | Adb69539 C. neofor |
| 24 | 1682 | 45.5 | 4738 | 10 ADB69178 | Adb69178 C. neofor |
| 25 | 1680 | 45.4 | 1969 | 8 ABT18375 | Abt18375 Aspergill |
| 26 | 1273 | 34.4 | 1664 | 4 AAK94164 | Aak94164 Human ful |
| 27 | 1273 | 34.4 | 1664 | 12 ADL30661 | Adl30661 Full leng |
| 28 | 1241.5 | 33.6 | 1209 | 4 AAH33264 | Aah33264 Human col |
| 29 | 1241.5 | 33.6 | 1209 | 6 ABL89850 | Ab189850 Human pol |
| 30 | 1241.5 | 33.6 | 1543 | 2 AAX85055 | Aax85055 Human sec |
| 31 | 1241.5 | 33.6 | 1543 | 8 ACD18981 | Ac18981 Novel hum |
| 32 | 1241.5 | 33.6 | 1543 | 12 ADG78372 | Adg78372 Human sec |
| 33 | 1241.5 | 33.6 | 1543 | 12 ADN60563 | Adn60563 Human sec |
| 34 | 890.5 | 24.1 | 2510 | 5 AAF93772 | Aaf93772 Human cdn |
| 35 | 890.5 | 24.1 | 2546 | 2 AAV44866 | Aav44866 Clone CTS |
| 36 | 890.5 | 24.1 | 2546 | 5 AAF98463 | Aaf98463 Human cdn |
| 37 | 880.5 | 23.8 | 2537 | 4 AAD08289 | Aad08289 Human cdn |
| 38 | 878.5 | 23.8 | 2284 | 4 AAH18021 | Aah18021 Human cdn |
| 39 | 875.5 | 23.7 | 2547 | 4 AAD08315 | Aad08315 Human sec |
| 40 | 875.5 | 23.7 | 2660 | 6 ABQ4750 | Abq4750 Human ova |
| 41 | 870.5 | 22.2 | 1114 | 4 AAH99794 | Aah99794 Human pro |
| 42 | 805 | 21.8 | 487 | 3 AAAL6208 | Aal6208 Human col |
| 43 | 802 | 21.7 | 455 | 6 ABL37867 | Ab137867 Human col |
| 44 | 793.5 | 21.5 | 787 | 4 AAH07526 | Aah07526 Human cdn |
| 45 | 792 | 21.4 | 440 | 6 ABK27622 | Abk27622 Human col |

ALIGNMENTS

RESULT 1

ADD94793
ID ADD94793 standard; DNA; 2472 BP.

XX AC ADD94793;

XX DT 29-JAN-2004 (first entry)

XX DE Human ITM1 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;

major histocompatibility complex; human leukocyte antigen; HLA; cytosolic; immunosuppressive; antisense therapy; gene therapy; cancer;

lung cancer; intestine cancer; sarcoma; prostate cancer;

testicular cancer; breast cancer; melanomas; pancreatic cancer;

haematological cancer; immune response; lymphoid cell proliferation;

autoimmune disease; transplant rejection; SIMP-derived peptide; human;

gene; ds; ITM1.

XX OS Homo sapiens.

XX PN WO2003054008-A2.

XX PD 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CR001967.

XX PR 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.

Db 1607 ATATTGATGATCCGAGAGCATATATTGGCTCGTCATATACCTCAGAGGATGG 1666
 Qy 521 LysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaSerThrIle 540
 Db 1657 AAGGTCAATGCTCGTGGGATATTGGCTATACAGATACAGATATGCAACACGACAAAT 1726
 Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1727 TTAGTGGACATAACACATGGAATAATACCATATTTCTGAGTAGGAGGCAATGGG 1786
 Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1787 TCCACAGAGGAAAGGCTATGAGATCATGAGGAGCTGATGTCAGTATGCTGGTC 1846
 Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 Db 1847 ATTTTGGAGGCTCACTGGGTATCTCTGATGATATCAACAAGTTCTTTGGATGGTC 1906
 Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrPro 620
 Db 1907 CGGATTGGAGGAGCAGATACAGGCAACATATCAAGGAGAAATGACTATTATACCTCA 1966
 Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1967 ACTGGGAGTTCCGTTGGGACCGGAGGTTCTCCAGTGTCTCAACTGCCTCATGTAC 2026
 Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAGCAAGCGTCTCCAGGCTTT 2086
 Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2087 GACCGTGTCCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGTCTCGGAGAA 2146
 Qy 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2147 GGCTATACACAGAACATGCTGCTGGTCAGGATATACAAAGTAAAGGACCTGGATAATCGA 2206
 Qy 701 GlyLeuSerArgThr 705
 Db 2207 GCTTGTCAAGGACA 2221
 RESULT 2
 ADH28844
 ID ADH28844 standard; DNA; 2472 BP.
 XX
 AC ADH28844;
 XX
 DT 11-MAR-2004 (first entry)
 DE Human chronic myelogenous leukaemia (CML) gene marker #112.
 XX
 KW ds; chronic phase chronic myelogenous leukaemia; CP-CML;
 KW blast crisis CML; BC-CML; human; chronic myelogenous leukaemia;
 KW gene marker.
 XX
 OS Homo sapiens.
 XX
 FN US2003104426-A1.
 XX
 PD 05-JUN-2003.
 XX
 PF 14-JUN-2002; 2002US-00171581.
 XX
 PR 18-JUN-2001; 2001US-0298914P.
 XX
 PA (LINS//) LINSLEY P S.
 PA (MAOM//) MAO M.
 PA (DAIH//) DAI H.
 PA (HEY//) HE Y.
 PA (RADI//) RADICH J P.
 XX
 PI Linsley PS, Mao M, Dai H, He Y, Radich JP;

XX
 DR
 XX
 XX
 PT
 PT
 PT
 XX
 PS
 XX
 CC
 CC
 CC
 CC
 CC
 CC
 XX
 SQ

WPI; 2003-787046/74.

Classifying cell sample as chronic phase chronic myelogenous leukemia or blast crisis chronic myelogenous leukemia by detecting difference in expression of genes corresponding to the markers such as X15415, U99436.

Disclosure; SEQ ID NO 112; 31pp; English.

The invention relates to a method of classifying a cell sample as chronic phase chronic myelogenous leukaemia (CP-CML) or blast crisis CML (BC-CML). The method is useful for classifying a sample as CP-CML or BC-CML. The present sequence represents a human chronic myelogenous leukaemia (CML) gene marker used to distinguish blast crisis CML from chronic phase CML.

Sequence 2472 BP; 566 A; 568 C; 583 G; 755 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2472
 Score: 3637.00 Matches: 705
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Ds: 10 Gaps: 0

US-10-028-384-12 (1-705) x ADH28844 (1-2472)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 107 ATGACTAAGTTTGGATTTTTCGATTTCTCTATGAGAAGCAGGACACACTTTTGAAGCTT 166
 Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 167 CTCTATTCGTCAATGGCTGCTGTATTATCTCTCCACTCGTCTGTTGCTGCTCTGAGA 226
 Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 227 TTTGAAAGTGTATTCATGATGAGTTTGATCCGCTACTTTAATTATCGGACTACCAAGTTCCTG 286
 Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
 Db 287 GCTGAGGAGGGGTTTATAAATTCATTAATCTGGTTGATGACCGAGCCTGGTACCTTTG 346
 Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 347 GGACGAATCATGAGGAGAACAAATTTACCAGAGTTTAATGATCACTCTCTGCAATCTAC 406
 Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 407 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAAATGCTGTGTGTCTGAGCC 466
 Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 467 CCT 526
 Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 527 GCAGGGCTGGGCTTCTGTCTGCTGCATGATTTCTGTAGTTCCTGGAATATCTCCCA 586
 Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 587 TCTGTGGCTGGCTCTATGATAATGAAGGATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 646
 Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 Db 647 TACATGTGGATCAAGGAGTAAAGATGTGGTTCATCTGTGGGAGCTAAGTGTGCTCTT 706
 Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro 220
 Db 707 GCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTCTCTGATCACTCACTAATTAAT 766
 Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240

Db 767 CTCCACGCTCTCGTCTGATGCTCACAGGCGGTTTCTCTCACGGATCTATGTGCGCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 Db 827 TGTACTGTTTACTGCTGGTGGTACTATATCTTCTAGGCAGATCTCTTGTGGTTCAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 Db 887 CCTCTCTCTTCTCATCAGACACATGGCAGGCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 946
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 947 GCCTTTGGTGGTACCTGGCAGCAGATTTGAATCCACACAAATTTGAAGTCTTTTCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuValGlyAlaLeuLeuMetLeuThr 320
 Db 1007 AGCGTCATCTCTCTGGTAGGCTTTGTCTCTCACCGTGGAGCTCTCTCATGTCTGACA 1066
 QY 321 GlyLysIleSerProTcPThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 1067 GGAATAATACTCCCTGGACGGGGCTTTCTACTCACTGCTGGATCCCTCTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 Db 1127 AACAAACATCCCATCATTTGCTGTCTGAGCATCAGCCACCAACCTGTCTCTCATAC 1186
 QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1187 TATTTTGAAGCTGAGCTCTCTCTCTCATGTTTCCAGTTGGCTCTTATTAAGCTTTTAC 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1247 AACCTGTCTGATGCCGAGATTTTATCATCATGATGGTGTGACCAAGCATGTACTTTTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1307 GCTGTATGTTGGTCTTAATGCTATGTTGGACCTGTTATGAGCATTTCTCTGGCAT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1367 GGAGTCTCCAGAGTCTGCTCCACATACATGAAGAATCTGACATAGTCCGCCACAG 1426
 QY 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 Db 1427 AAGAGCAAGAGCAGCAGGATTCACCTACCTATTAAGATTGAAGTGGCAGTGGGATG 1486
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 Db 1487 ATACTGTGATGGCTTTCTTCTCATCACCTACCATCCTTTTCACTCAACCTGGTGACCA 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyLysPglySerArgile 500
 Db 1547 GAGGCTTACTTCTCTCGTCCATTTGATCTATCTGCCCGTGGTGGGATGGCAGTAGATC 1606
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 Db 1607 ATATTGTGATGACTCCGAGAGCATATATTGGCTTTCTGTCATAATCTCCAGAGGATCG 1666
 QY 521 LysValMetSerTrpTcPAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 Db 1667 AAGTCTATGCTCTGGTGGATTTAGCTATCAGATTACAGCTATGGCAACCGACAAAT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 Db 1727 TTAGTGACATAAACACATGAAATAATACCATATTCTCGAGTAGGCGCAGCAATGGCG 1786
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 Db 1787 TCCACAGAGGAAAGCCTATGAGATCATGAGGAGCTCGATGTGAGCTATGTGCTGCTC 1846
 QY 581 IlePheGlyLeuLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

Db 1847 ATTTTGGAGGCTCACCTGGGTATTCTCTGATGATATCAACAGTTTCTTTGGATGGTC 1906
 QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 Db 1907 CGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1967 ACTGGGAGTTCCTGGTGGACCGTGAAGGTTCTCCAGTGTGCTCAACTGCCCTCATGTAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2027 AAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAACCAAGCGTCTCTCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 Db 2087 GACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCTGGAGAA 2146
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2147 GGCATATACACAGAACATTGGCTGTCAGGATATACAAAGGTAAAGGACCTGGATATATCGA 2206
 QY 701 GlyLeuSerArgThr 705
 Db 2207 GGCCTGTCAAGGACA 2221
 RESULT 3
 ID ADD94791 standard; DNA; 3094 BP.
 XX ADD94791;
 AC ADD94791;
 DT 29-JAN-2004 (first entry)
 XX Mouse ITM1 gene sequence.
 DE source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytotoxic; immunosuppressive; antitense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; gene; ds; ITM1.
 XX Mus musculus.
 OS WO2003054008-A2.
 XX 03-JUL-2003.
 XX 18-DEC-2002; 2002WO-CA001967.
 XX 20-DEC-2001; 2001US-00028384.
 XX (COMP-) COMPATIGENE INC.
 XX Perreault C, McBride K;
 XX WPI; 2003-559122/52.
 XX P-PSDB; ADD94792.
 XX PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 XX acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 XX or breast cancer, or for suppressing an immune response in an autoimmune
 XX disease.
 XX Disclosure; SEQ ID NO 9; 66pp; English.
 PS This invention relates to a novel isolated or purified human protein.
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein

CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC cancer, melanomas, pancreatic cancer, testicular cancer, breast
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the mouse Irf1 gene which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.

XX SQ Sequence 3094 BP; 758 A; 651 C; 703 G; 982 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 3094
 Score: 3633.00 Matches: 694
 Percent Similarity: 98.87% Conservatives: 3
 Best Local Similarity: 98.44% Mismatches: 8
 Query Match: 98.27% Indels: 0
 DB: 10 Gaps: 0

US-10-028-384-12 (1-705) x ADD94791 (1-3094)

| | | | |
|----|-----|---|-----|
| QY | 1 | MetThrIysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu | 20 |
| DB | 112 | ATGACATAAGCTGGATTTTCGGATTGCTCTATGAGAGGAGGACACATCTCTAAGCTT | 171 |
| QY | 21 | LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg | 40 |
| DB | 172 | CTCATCTGCTGATGGCTGCTGTATCTTTTCTACTGCTCTTTGCTGCTGAGA | 231 |
| QY | 41 | PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu | 60 |
| DB | 232 | TTTGAAGTGTGATCCATGAGTGTGATCCGCTACCTTTAATATCGGACTACCCGGTTCTG | 291 |
| QY | 61 | AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTyrTrpProLeu | 80 |
| DB | 292 | GCTGAGGAGGGGTTTATTAATTCATAACTGTTTGTATGACCGGGCTTGTACCTTGTG | 351 |
| QY | 81 | GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr | 100 |
| DB | 352 | GCCCAATCATTTGGAGGAGCAATTTACCCAGGTTTAAATGATCACTTCTGTCGCAATCTAC | 411 |
| QY | 101 | HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla | 120 |
| DB | 412 | CATGTAATCCATTTCTTCCATATCATATGACATTCGGATGCTGCTGTTTCTGCTGCC | 471 |
| QY | 121 | ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp | 140 |
| DB | 472 | CCACTTTTCTCTCTTCCACCACCATGCTTACGTACCACTTACCAAGAGCTCAAGGAT | 531 |
| QY | 141 | AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyrIleSerArg | 160 |
| DB | 532 | GCAGGAGCTGGCTTCTGCTGCTGCCATGATGCTGTAGTTCCTGGATATATTTCTCGA | 591 |
| QY | 161 | SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr | 180 |
| DB | 592 | TCGTAGTGGCTGCTATGATAATGAAGAAATTCATCTTTTTCATGCTGCTTACTTAC | 651 |
| QY | 181 | TyrMetTrpIleLeuAlaValIleThrGlySerIleCysTrpAlaAlaLysCysAlaLeu | 200 |
| DB | 652 | TACATGTGGATCAAGGAGTGAAGCTGTTCCATCTATTGGCTGGCAAGTGTGCCCTC | 711 |
| QY | 201 | AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuLeuPro | 220 |
| DB | 712 | GCTTATTTCTACATGGCTCTTCTATGGGAGGCTATGTTTCTGTGATCAACTTGAATCTCT | 771 |
| QY | 221 | LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr | 240 |

| | | | |
|----|------|---|------|
| DB | 772 | CTACATGCTCTGGTGCTAATGCTGACAGGCGCGTTTTTCTCACCGGATCTACGTAGCCTAC | 831 |
| QY | 241 | CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln | 260 |
| DB | 832 | TGTACTGTTTACTGCTGGGACCAATCTTCTTATGACGATTTCTTTGTTGTTCCAG | 891 |
| QY | 261 | ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis | 280 |
| DB | 892 | CCGCTGCTTTTCATCAGAACACATGGCAGCCCTTTGGAGTGTGTTGCTCTCTGTGACATCCAT | 951 |
| QY | 281 | AlaPheValAspTyrLeuAspGlySerLysLeuAsnProGlnPheGluValLeuPheArg | 300 |
| DB | 952 | GCTTCTGTAGATACCTGCGCAGCAGATGGAATCCACAGCAATTCGAAGTCTTTTCCGG | 1011 |
| QY | 301 | SerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetLeuThr | 320 |
| DB | 1012 | AGTGTATCTCTCTGCTGTTGGCTTTGCTCTCTCACTGTGGAGCTCTCTCATGTCAACA | 1071 |
| QY | 321 | GlyLysIleSerProTyrTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys | 340 |
| DB | 1072 | GGAAAAATTTCTCCCTGGACAGGCGGTTTCTACTCTCTGCTGGATCCCTTCTTATGCTAAG | 1131 |
| QY | 341 | AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr | 360 |
| DB | 1132 | AATAACATTCCTATTTATGATCTGTTCTGAGCAGCCACCAACCTGGTCTTCTCTAC | 1191 |
| QY | 361 | TyrPheAspLeuLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer | 380 |
| DB | 1192 | TATTTGATCTACAGCTCTCTGCTCTTCTCATGTTTCCAGTTGGCTCTTACTACTGCTTTAGC | 1251 |
| QY | 381 | AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer | 400 |
| DB | 1252 | AACCTGCTGATGCTCGGATTTTATCATCATGATGTTGGTGACAGCATGATCTTTTCA | 1311 |
| QY | 401 | AlaValMetValArgIleMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle | 420 |
| DB | 1312 | GCTGTATGTTGGCTCTAATGCTGTTATGCGACCTGTTATGTCGATCTTCTTCTGCAAT | 1371 |
| QY | 421 | GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys | 440 |
| DB | 1372 | GGTGTTTCCAGGTTGCTCCACATATATGAAAAATCTGGACATAAGTCCGCCAGACAAG | 1431 |
| QY | 441 | LysSerLysLysGlnGlnAspSerThrTyrProIleIleGluValAlaSerGlyMet | 460 |
| DB | 1432 | AAGAGCAGAGAGCAACAGGATCTTACTTACCTTATTAAGAAATGAGTGGCGAGTGGGATG | 1491 |
| QY | 461 | IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer | 480 |
| DB | 1492 | ATCTGTTGATGCTGCTTTTCTCATCACCTACACGTTTCTTCTGCTGCTGGGTGACCAT | 1551 |
| QY | 481 | GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle | 500 |
| DB | 1552 | GAGGCTATTTCTCTCCCTCCATGTTACTGCTGCTGGTGGGATGGCAGTAGGATC | 1611 |
| QY | 501 | IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla | 520 |
| DB | 1612 | ATTTTGTGATCTTCCGAGAGGATTAATTTGGCTCCGTCAACAATCTCCAGAGGATGCA | 1671 |
| QY | 521 | LysValMetSerTrpTrpAspTyrGlyThrGlnIleThrAlaMetAlaAsnArgThrIle | 540 |
| DB | 1672 | AAGTCATGTCATGGTGGGATTTATGGCTACCAATTTACTGCAATGGCAATTCGACAAAT | 1731 |
| QY | 541 | LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla | 560 |
| DB | 1732 | TTAGTGACATAACACATGGAATAATACCATATTTCTCGAGTAGGAGGCAATGGCA | 1791 |
| QY | 561 | SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal | 580 |
| DB | 1792 | TCCACAGAGAAAAAGCCTATGAAATCATGAGGAGCTTGTATGTCAGCTATGCTGTGTC | 1851 |
| QY | 581 | IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal | 600 |
| DB | 1852 | ATTTTGGAGGCTTACTGGGTATTTCTCGATGATATCAACAAGTTTCTTTTGGATGGTC | 1911 |

QY 601 ArgTleGlyGerThrAspThrGlyLysHisLysGluAsnAspThrPro 620
 Db 1912 CGGATTGGAGGAAACACAGACAGAGAGAGACACATTAAGGAGAGTACTATATCTCT 1971
 QY 621 ThrGlyGluPheArgValAspGlyGlySerProValLeuLeuAsnCysLeuMetTyr 640
 Db 1972 ACTGGGAATTCGGTGTGATCTGAGGGTCTCCGGTCTGCTCACTGCGCTTAATGAC 2031
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2032 AAAATGTGTACTACCGCTTTGGCGAGTCTACACAGAGCAAGCGTCCACAGGCTTT 2091
 QY 661 AspArgValArgAsnAlaGluLleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 Db 2092 GACCGTGTTCGAATGTGAGATGGTAATTAAGACTTTGAGCTTGATCTCTGGAGAA 2151
 QY 681 GlyTyrThrGluHisTyrLeuValArgLysTyrLysValLysAspLeuAspHarg 700
 Db 2152 GCGTATACACAGAACACTGGTAGTCAGGATATACAGGTAAGGACCTGGATATCGA 2211
 QY 701 GlyLeuSerArgThr 705
 Db 2212 GGCTGTGACAGGACA 2226
 RESULT 4
 ID ABV24502 standard; cDNA; 2760 BP.
 XX AC ABV24502;
 XX DT 16-SEP-2002 (first entry)
 XX DE Human prostate expression marker cDNA 24493.
 XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 XX KW pharmacogenomic marker; gene; ss.
 XX OS Homo sapiens.
 XX PN WO200160860-A2.
 XX PD 23-AUG-2001.
 XX PF 20-FEB-2001; 2001WO-US005171.
 XX PR 17-FEB-2000; 2000US-0183319P.
 XX PR 16-MAR-2000; 2000US-0189862P.
 XX PR 25-MAY-2000; 2000US-0207454P.
 XX PR 09-JUN-2000; 2000US-0211314P.
 XX PR 18-JUL-2000; 2000US-0219007P.
 XX PR 13-DEC-2000; 2000US-0255281P.
 XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX PI Schlegel R, Endege WO, Monahan JE;
 XX WP1; 2001-662795/76.
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX Claim 1; Page 4632-4633; 11750pp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;
 SQ Alignment Scores:
 Pred. No.: 0 Length: 2760
 Score: 3631.00 Matches: 698
 Percent Similarity: 98.87% Conservative: 1
 Best Local Similarity: 98.73% Mismatches: 6
 Query Match: 98.21% Indels: 2
 DB: 5 Gaps: 0
 US-10-028-384-12 (1-705) x ABV24502 (1-2760)
 QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
 Db 105 ATGACTTAAGTTTGGATTCTTTCGATTGTCTTATGAGAAGCAGGACACACTTTTGAAGCTT 164
 QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr 40
 Db 165 CTCAATCTCTCAATGGCTGCTGATATTCCTTCTCCACTCGCTCTGTTTGTCTGCTCTGAG 224
 QY 40 gPheGluSerValLleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLe 60
 Db 225 ATTGAAGTGTATCCATGAGTTTGTATCCGCTACTTTAATATATCGGACTACCAAGTTCCT 284
 QY 60 uAlaGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTrpProLe 80
 Db 285 GGTCTGAGGAGGGGTTTATAAATCCATACTGTTTGTATGACCGACCTGGTACCCCTTT 344
 QY 80 uGlyArgLleLleGlyThrLleTyrProGlyLeuMetLleThrSerAlaLleTyr 100
 Db 345 GGGAGCAATCATTTGAGGAAACAATTTACCAAGTTTAATGATACCTCTGCTGCAATCTA 404
 QY 100 rHisValLeuHis-PhePheHisLleThrLleAspLleAtgAsnValCysValPheLeuA 120
 Db 405 CCATGTACTCCATTTTTCACATCCACATCGAATTCGGAATGCTGTGTGTCTCTGG 464
 QY 120 laProLeuPheSerSerPheThrSerLleValThrTyrLeuLeuThrLysGluLeuLysA 140
 Db 465 CCCTCTCTCTCTCTCTCCATCCACCATCGTCAGCTACCACTTTACCAAGAGCTCAAGG 524
 QY 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetLleAlaValValProGlyTyrLleSerA 160
 Db 525 ATGCAGGGGCTGGGCTTCTTGTCTGCTGCCATGATTCGTAGTCTCTGGATATATCTCCC 584
 QY 160 rGSerValAlaGlySerTyrAspAsnGluGlyLleAlaLlePheCysMetLeuLeuThrT 180
 Db 585 GATCTGTGGCTGGCTCTCTATGATATATGAAGGATGTCATCTTTTGTGCTACTACCT 644
 QY 180 yTyrMetTrpLysAlaValLysThrGlySerLleCysTrpAlaAlaLysCysAlaL 200
 Db 645 ACTACATGTGGATCAAGGAGGAGTAAAGACTGTTTCACTGTGTGGGAGCTAAGTGTGCC 704
 QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuP 220
 Db 705 TTGCTTATTTCTACATGCTCTGCTCATGGGAGGTTATGTCTTCTGATCAACTTAATTC 764
 QY 220 roleuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgLleTyrValAla 240
 Db 765 CTTCTCACGCTCTCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGCT 824
 QY 240 yTyrCysThrValTyrCysLeuGlyThrLleLeuSerArgGlnLleSerPheValGlyPhe 260
 Db 825 ACTGTACTGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
 QY 260 laProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuLysLeuH 280
 Db 885 AGCTGTCTCTCTATCAGAGACATGGGAGGCTTTGGGCTTTTGGTCTCTGCGAGATCC 944

QY 280 isAlaPheValAspTyrLeuArgSerIysLeuAsnProGlnGlnPheGluValLeuPheA 300
 DB 945 ATGCGCTTTGGGATTAACCTGGCGAGCAAGTTGAATCCACAACTTTGAAGTTCTTTCC 1004
 QY 300 rGSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuT 320
 DB 1005 GGAGGGGTATCTCTCTGGAGCTTTGCTTCTCACCGTGGAGCTCTCTCATGCTGA 1064
 QY 320 hrGlyIleSerProThrTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaL 340
 DB 1065 CAGGAAAAATATCTCCCTGGACGGGGCGTTTCTACTCGCTCGCTGGATCCCTCTTATGCTA 1124
 QY 340 ysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerT 360
 DB 1125 AGAACACATATCCCAATCATTTGCTTCTGTCTGAGCATCAGCCACCAACCTGGTCTCAT 1184
 QY 360 yTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheS 380
 DB 1185 ACTATTTTACCTGAGCTCTCTGCTCTTCACTGTTTCCAGTTGGCTCTATTACTGCTTTA 1244
 QY 380 erAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheS 400
 DB 1245 GCAACCTGCTGATGCCGGATTTTATCATCATGATGTTGACCGACGATGATCTTTT 1304
 QY 400 erAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyI 420
 DB 1305 CAGCTGTAATGCTGCTCTAAATGCTAGTGTGGCACCTGTTATGTGCATTTCTCTGGCA 1364
 QY 420 leGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspL 440
 DB 1365 TTGGAGTCTCCAGGTGCTGCTCCACATACATGATGAGATCTGGACATAGTCTGCCAGCA 1424
 QY 440 ysLysSerIleLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyM 460
 DB 1425 AGAAGAGCAAGCAAGCAAGATCCACCTACCTCTTAAAGATGAAGTGGCAAGTGGGA 1484
 QY 460 erIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValTrS 480
 DB 1485 TGATACTGCTATGGCTTTCTTCTCATCCTACCTACCTTTTCATCAACCTGGGTGACCA 1544
 QY 480 erGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgI 500
 DB 1545 GTGAGGCTACTCTCTCCGCTCCATTTCTGTCTGCGCGTGTGGGGATGGCAGTAGGA 1604
 QY 500 leilePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspA 520
 DB 1605 TCATATTTGATGACTTCCGAGAGCATATATTTGGCTTCTGTCATATATCTCCAGAGATG 1664
 QY 520 laLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrI 540
 DB 1665 CGAAGGTATGCTCTGCTGGTGGATTTATGGCTATCAGATTACAGCTATGGCAACCGAACA 1724
 QY 540 leLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetA 560
 DB 1725 TTTTATGTGACATATPACACATGGAATATATACCCATTTTCTCGATAGGCGACGATGG 1784
 QY 560 laSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuV 580
 DB 1785 CGTCCACAGAGAAAAGCTTATGAGATCATGAGGAGCTCGATGTCTGAGTATGTCTGCTG 1844
 QY 580 aLilePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetV 600
 DB 1845 TCATTTTGGAGGCTCTCTGGTATTTCTCTGTATGATATCAACAGTTTCTTTTGGATGG 1904
 QY 600 alArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrP 620
 DB 1905 TCCGATTTGGAGGAGCAGATACAGGCAACATATCAAGGAGATGATGATATTATCTC 1964
 QY 620 roThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetT 640
 DB 1965 CAACTGGGAGATTCGCTGTGACCGTGAGGTTCTCCAGTGTCTGCTCAATGCTGCTATGT 2024
 QY 640 y-LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyP 660

DB 2025 ACAAGATGTGTACTATCGCTTTGGACAGGTTTACACAGAACCAAGCGCTCTCCAGGCT 2084
 QY 660 heAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluG 680
 DB 2085 TTGACCGTGTCCGAAATGCTGAGATTGGGATTAAGACTTTGAGCTTGATGCTCTGGAGG 2144
 QY 680 luGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnA 700
 DB 2145 AAGCATATACACAGAACATTGGCTGTGTCAGGATATACAGGTAAAGGACCTGGATAATC 2204
 QY 700 rGlyLeuSerArgThr 705
 DB 2205 GAGGCTTGTCAAGACA 2221
 AC ABL02795;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2867.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmacological; gene; ss.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US009231.
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR MPI; 2001-656860/75.
 DR P-PSDB; ABB58692.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PS Claim 1; SEQ ID NO 2867; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 SQ Sequence 2855 BP; 661 A; 750 C; 724 G; 720 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.21e-284 Length: 2855
 Score: 2903.50 Matches: 536
 Percent Similarity: 87.02% Conservative: 74
 Best Local Similarity: 76.46% Mismatches: 84
 Query Match: 78.54% Indels: 7
 Gaps: 4

RESULT 6
 ABL02794
 ID ABL02794 standard, cDNA; 6153 BP.
 XX
 AC ABL02794;
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2864.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 XX
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEXE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB58691.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 2864; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU130511), expressed DNA
 CC sequences (ABU1840-ABU16175) and the encoded proteins (ABB57737,
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6153 BP; 1692 A; 1334 C; 1299 G; 1828 T; 0 U; 0 Other;

Alignment Scores:

| Pred. No.: | 1-62e-253 | Length: | 5153 |
|------------------------|-----------|---------------|------|
| Score: | 2602.50 | Matches: | 524 |
| Percent Similarity: | 60.24% | Conservative: | 70 |
| Best Local Similarity: | 53.14% | Mismatches: | 79 |
| Query Match: | 70.39% | Indels: | 315 |
| DB: | 4 | Gaps: | 5 |

US-10-028-384-12 (1-705) x ABL02794 (1-6153)

| | | |
|----|---|------|
| QY | 30 SerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHisGluPheAsp | 49 |
| DB | 1890 GCATTGGCACACGCTTCTCTGTCTGCGATTCGAAACGCGTAATCCATGAGTTTCGAT | 1949 |
| QY | 50 ProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGlyPheTyrLysPheHis | 69 |
| DB | 1950 CGGTACTTCACTACCGCACACCGGGTTCTGGCGGAGGCGGCTTTACAGTTCCAC | 2009 |
| QY | 70 AsnTrpPheAspArgAlaTrpTyrProLeuGlyArgIleGlyGlyThrIleTyr | 89 |
| DB | 2010 AACTGGTTTCGATGACCGCGCTGGTATCCCTGGCGGCGATCATCGCGGCACCATCTAT | 2069 |
| QY | 90 ProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHisPhePheHisIleThr | 109 |

| | | |
|----|--|------|
| DB | 2070 CCGGGCTGATGCTCACCTCGCGGCGCTGTACCGCTGTGTGCTGCTCAATGTGACC | 2129 |
| QY | 110 IleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerSerPheThrSerIle | 129 |
| DB | 2130 ATGCACATACGGAACGTTGCTGCTTCTCTGGCGCCCTTCTTCTCTCGCTGACCGCTG | 2189 |
| QY | 130 ValThrTyrLeuLeuThrLysGluLeu- | 138 |
| DB | 2190 GTGACCTACGCCCTCACAAAGGAGATACACGTAAAGCCACCGTCCACCGTCTTCTTACTTT | 2249 |
| QY | 138 | 138 |
| DB | 2250 TTTTATATATCTTCGAGGTTAAATTCGACGCTGTCTTATGATCCCTTTATCC | 2309 |
| QY | 139 | 156 |
| DB | 2310 CTTTATTAGAGC-ACCTGGAGCTGGACTGGTGGCGCGCTTGTATATCCATCGTTCCCGG | 2368 |
| QY | 157 TyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMet | 176 |
| DB | 2369 TATATCTCGATCCGTTGGCGGATCGTAGCAATGAAGGCATCGCCATTTTCTCATG | 2428 |
| QY | 177 LeuLeuThrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAla | 196 |
| DB | 2429 CTCTTACCTACTATTGTGATCAAGCGGTAAAGACGGGCACGATCTTTTGGTCGGCT | 2488 |
| QY | 197 LysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIle | 216 |
| DB | 2489 ATGTCGCATTGGCTTACTTCTATATGCTCTCTCTGGGGTGGCTATGCTTCTTGAT | 2548 |
| QY | 217 AsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIle | 236 |
| DB | 2549 AACCTAATCCCGTGCACGCTGGCGGTGATCATCACGGACGTTTCTCGCACAGATC | 2608 |
| QY | 237 TyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPhe | 256 |
| DB | 2609 TACATGATACACGACGCTATCTGCTGGCGACCATCTCTGCTGATGAGATCTCGTT | 2668 |
| QY | 257 ValGlyPheGlnProValLeuSerSerGluHisMetAlaGly | 270 |
| DB | 2669 GTGGGATTCACACCCATCCAGAGCTCGAACAACATGCT-GGTGAGTGTAACTACTTTGTG | 2727 |
| QY | 270 | 270 |
| DB | 2728 TAATGCTGTGAGTGTAGCATGTGTGATCTGTGTGATGATACATTTTAAATTTTGA | 2787 |
| QY | 270 | 270 |
| DB | 2788 TAGGACGAGTATGAACATATGATGATACCAAAACCTAGTGAATGGTCACTAAAGAA | 2847 |
| QY | 270 | 270 |
| DB | 2848 CTCTAATTAAGAATTAATCTGCTATATCGAAAGAGGAGGTGAATCGTATTTTCTC | 2907 |
| QY | 271 | 287 |
| DB | 2908 TTGCAGGCACCTGGGAACCTTTGGCTGTGCCAGATTCACGCTTTCGCTGACATCTGCGC | 2967 |
| QY | 288 SerLysLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGly | 307 |
| DB | 2968 TCGCGCATTCACAGGATCACTTCGATCTGCTCTTCAAGACGTTGGTTTCCAGTGTGTTG | 3027 |
| QY | 308 PheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProTrpThr | 327 |
| DB | 3028 ACTGTGTGTTCGTCGGTGACCTGCTCACGCTTACCGGGAAGTCTCCCGTGGACC | 3087 |
| QY | 328 GlyArgPheTyrSerLeuLeuAspProSerTyrAlaIleAsnAsnIleIleAla | 347 |
| DB | 3088 GGCAGATTTTACTCGCTACTAGATCCATCTATGCCAAGATCACATCCATCATCGCC | 3147 |
| QY | 348 SerValSerGluHisGlnProThrThrTrpSerSerTyrTyrPheAsp | 363 |

Db 3148 TCCGATCGGAGCATCAGCCTACATCCTGCTGCTCTTCTATTTGATCTGCAGTAGGT 3207
 QY 363 ----- 363
 Db 3208 AAACCACTATCTCGTAGGAATCTGTAGATGCTCTCACTTAACAGTGTGCTTATC 3267
 QY 363 ----- 363
 Db 3268 ATATCTATTTTGTGTTAAATAATCCTTACTACAAACAAATTCGGTTTATTTTAAACCTAT 3327
 QY 363 ----- 363
 Db 3328 CCTAATTACACATGTTCTCGGAGGATGAGAACTTAACACGATGGAATGATGCGCC 3387
 QY 363 ----- 363
 Db 3388 TTAATGCAATTGAGTATTACCTTGACATTTGATCGCTGAGCAGTTAAATAACATGACC 3447
 QY 364 ----- 364
 Db 3448 CCTTTTACATAACATCCAAATATATAGTTAATTCATTCGCTTTTTCACAAATTCCTTT 3507
 QY 364 uGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAs 384
 Db 3508 GCAGATCCTGGTCTCTCTTCCCGCTGCGCTTTATTTCTGCTTTTCCAAAGCTGACGGA 3567
 QY 384 pAlaArgilePheileileMetTyrGlyValThrSerMetTyrPheSerAlaValMetVa 404
 Db 3568 CTCGAACATCTTATATTTTGTATGGCTCACCAGTATTATTTTCGCGGAGTATGATGT 3627
 QY 404 lArgLeuMetLeuValLeuAlaProValMetSerileLeuSerGlyileGlyValSerGl 424
 Db 3628 GCGTCTAATGCTGCTCTCGCGCGCTGATGTGTGCTATCCGAAATGGCAATTTCGCA 3687
 QY 424 nValLeuSerThrTyrMetLysAsnLeuAspilleSerArgProAspLysLysSer---- 442
 Db 3688 TCTCTGGCCAAATATATCAAGAGCGTTGATCGCGAAAGCTCATCGAGCGAGGAGTAGA 3747
 QY 443 -----LysLysGlnGlnAspSerThrTyrProilleLysilleGluValAl 457
 Db 3748 AAGCAAGCGGAGCAGCAAGAGTGGACACAGCAGCAGCGGAGCGCTGAAGAGCGAGGTCGC 3807
 QY 457 aserGlyWetileLeuValMetAlaPhePheLeuileThrTyrThrPheHisSerThrTr 477
 Db 3808 CATTTGGATTCGTGGGTCTCATCTCTGATGCTTATAGTGTACAGCTGCACCTGCACCTG 3867
 QY 477 pValThrSerGluAlaTyrSerSerProSerileValLeuSerAlaArgGlyGlyAspGl 497
 Db 3868 GGTCACTCGAGGCGCTTACTCTCGCCAGTATTTGTTGAGTGCAGGTGCACGATGG 3927
 QY 497 YserArgileilePheAspAspPheArgGluAlaTyrTyrTrpLeuAlaGHisAsnThrPr 517
 Db 3928 CGGCCGGATCATTTTCGATGACTTCGCGAGGCGCTTACTGCTGCTGCAGATCAACACTCC 3987
 QY 517 oGlu----- 518
 Db 3988 GGAGGTGAATATGAAGTTTTGATTGCCCATTTTGTGCGATGCTCTCGCGAGTTAGCCATAG 4047
 QY 518 ----- 518
 Db 4048 AAATCAGAAACAGATTTGCTAATGGAATAATTGTGGGCTTTTAAAGATCGGAATTTCA 4107
 QY 518 ----- 518
 Db 4108 TCTGGCTAAGTCGGCAACACAGAGTTTCGATCTTTAATGATGACTACATTTGTAATTTCT 4167
 QY 518 ----- 518
 Db 4168 TGATAGGAGAGTATGAACATAATGATGATTAATAACCAAGCTGACAAACAGAGATA 4227
 QY 518 ----- 518
 Db 4228 TCCCTTTTGTGCAAACTTTTGTGTTAAATGCTGCGACCTGACCGCTTTCGTAAATGATTTT 4287

QY 519 ----AspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnileThrAlaMetAlaAs 537
 Db 4288 TCAGGAGGCTCGCAATATGCTCTGCTGGACTACGGCTACAGATAACGCCCATGGCCAA 4347
 QY 537 nArgThrileLeuValAspAsnAsnThrTrpAsnAsnThrHisileSerArgValGlyGl 557
 Db 4348 TCGGACGATATTGGTGGATAACAATCTTGGAAACAACACATATATCGCGCGTGGCCA 4407
 QY 557 nAlaMetAlaSerThrGluGluLysAlaTyrGluileMetArgGluLeuAspValSerTy 577
 Db 4408 GCGATGGCTCTTCGGAGGAGAAAGCCTACAGATAATGAGGAACTGGATGTGGACTA 4467
 QY 577 rValleuValillePheGlyGlyLeuThrGlyTyrSerSerAspAspileAsnLysPheLe 597
 Db 4468 CGTTCTCGTGATTTTCGGAGGCGCTCACGGCTACTCATCGACATATCAACAGTTCTCT 4527
 QY 597 uTrpMetValArgileGlyGlySerThrAspThrGlyLysHisileLysGluAsnAspTy 617
 Db 4528 GTGGATGTGGCAATTGGCGCAGCAGCATCGTGTGGCAGCATCCGCGAAAGGACTA 4587
 QY 617 rTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValleuLeuAsnCy 637
 Db 4588 CTATCGCGCAACGAGAGTTCCGAGTGGACAAGAGGGGCTCACCCACACTGCTCAATTG 4647
 QY 637 sLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPr 657
 Db 4648 TTTGATGTACAAATGTCTCTACTATCGCTTTGGGCAAAATGTACACGAGGTGGCAAGC 4707
 QY 657 oProGlyPheAspArgValArgAsnAlaGluileGlyAsnLysAspPheGluLeuAspVa 677
 Db 4708 CCAGGCGCTACGATCGAGTTCGTGCGCGCGAGATCGGCAACAGGACTTTGAATGGATGT 4767
 QY 677 lLeuGluGluGlyTyrThrThrGluHisTrpLeuValArgileTyrLysValLysAspLe 697
 Db 4768 CATGGAGGAGGCGGTACACCCAGGAGCTGCTGCTGGTGGCATCTTACAAGGTTAAGGATCT 4827
 QY 697 uAspAsnArgGlyLeu 702
 Db 4828 GCCGAATCGTGGAGTC 4843
 RESULT 7
 ADP28508
 ID ADP28508 standard; DNA; 1371 BP.
 XX
 AC ADP28508;
 XX
 DT 12-AUG-2004 (first entry)
 XX
 DE Human secreted protein encoding sequence SEQ ID #506.
 XX
 KW Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
 XX cancer; inflammatory; immune; ds; human secreted protein.
 OS Homo sapiens.
 XX
 FN WO2004035732-A2.
 XX
 PD 29-APR-2004.
 XX
 PF 28-AUG-2003; 2003WO-US026780.
 XX
 PR 29-AUG-2002; 2002US-0406576P.
 PR 29-AUG-2002; 2002US-0406579P.
 PR 29-AUG-2002; 2002US-0406585P.
 PR 29-AUG-2002; 2002US-0406588P.
 PR 29-AUG-2002; 2002US-0406608P.
 PR 29-AUG-2002; 2002US-0406611P.
 PR 29-AUG-2002; 2002US-0406612P.
 PR 29-AUG-2002; 2002US-0406614P.
 PR 29-AUG-2002; 2002US-0406642P.
 PR 29-AUG-2002; 2002US-0406646P.

Db CCTGTCCTTTTCATCAGAGCATGGGAGCCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 840
 Qy 281 AlapheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
 Db 841 GCCCTTTGGGAATACCTCGCGAGCAAGTTGAATCCCAACAATTTGAAGTCTTTTCCGG 900
 Qy 301 SerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetLeuThr 320
 Db 901 AGCGTCATCTCTCGTAGGCTTTGCTCTCTCACCGTGGAGCTCTCTCTCATGCTG-- 957
 Qy 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 Db 958 GGAAAAATATCTCCCTCGAGCGGGCGTTCATCTCTGCTGATCCCTCTTATGCTAAG 1017
 Qy 341 AsnAsnIleProIleLeuAlaSerValSerGluHisGlnProThrThrThrPheSerTyr 360
 Db 1018 AACAAATCCCATCATTTGCTCTGTGCTGTGAGCATCAGCCCAACCTGGTCCCTCATAC 1077
 Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 Db 1078 TATTTTGACCTGCGAGCTCTCGTCTTCATGTTTCCAGTTGGGCTCTTATTAAGCTTAC 1137
 Qy 381 AsnLeuSerAspAlaArgIlePheLeuMetTyrGlyValThrSerMetTyrPheSer 400
 Db 1138 AACCTGCTGATGATCCCGGATTTTATCATCATGATGTTGTCAGCAGCATGACTTTTCA 1197
 Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 Db 1198 GCTGTAATGGTGGCTCTAAAGCTAGTGTGGACCTGTTATGTCATTTCTCTC--ATT 1255
 Qy 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 Db 1256 GGAGTCTCCAGGTGCTGTCCACATCATGAAGAATCTGGACATAAGTGTCCAGACAG 1315
 Qy 441 LysSerLysLysGlnGlnAsnSerThrTyrProIleLysIleGluValIleSerGly 459
 Db 1316 AAGAGCAAGAAAA--CAGGATTCACCTACCTCTATTAAAGATGAAGTGGCAAGTGGG 1370

RESULT 8

ADD94783
 ID ADD94783 standard; cDNA; 2481 BP.

XX AC ADD94783;
 XX DT 29-JAN-2004 (first entry)
 XX DE Human SIMP cDNA sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cystostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanoma; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; human;
 KW gene; ss.

XX OS Homo sapiens.
 XX FN WO20003054008-A2.
 XX PD 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CA001967.
 XX PR 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.
 XX PI Perreault C, McBride K;
 XX DR WPI; 2003-559122/52.

P-PSDB; ADD94784.

XX New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins, useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.

XX Claim 6; SEQ ID NO 1; 66pp; English.

XX This invention relates to a novel isolated or purified human protein, MHC-
 CC termed source of immunodominant major histocompatibility complex (MHC)-
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytostatic or immunosuppressive activity or provide sequences useful for
 CC anticancer therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanoma, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the human SIMP cDNA which is related to the
 CC invention. Note: This sequence does not appear in the specification but
 CC was obtained by the indexer from GenBank.

XX SQ Sequence 2481 BP; 683 A; 503 C; 577 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,366-218 Length: 2481
 Score: 2247.00 Matches: 427
 Percent Similarity: 73.9% Conservative: 102
 Best Local Similarity: 59.7% Mismatches: 150
 Query Match: 60.7% Indels: 36
 DB: 10 Gaps: 7

US-10-028-384-12 (1-705) x ADD94783 (1-2481)

Qy 16 ThrLeuLeuLysLeuLeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 Db 208 TCGCTTCCTCTCTCCACCATCTCTCTCTCGGCGCTTGCCTTTCAGCTCGCGCTC 267
 Qy 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 Db 268 TTGCGCGCTCATCCGCTTCGAAGCATCATCCAGGTTCCAGCCGGTGTAACTATAGA 327
 Qy 56 ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgArg 75
 Db 328 TCAACACATCATCTTGCATCTCATGGTTCATGAAATTTTAAATGGTTTGAAGAAGA 387
 Qy 76 AlaTyrTyrProLeuGlyArgIleAlaGlyGlyThrIleTyrProGlyLeuMetIleThr 95
 Db 388 GCATGGTATCCACTAGGAAGAATAGTAGTGGTGTACTGTTTACCAGGGTTGATGATAACC 447
 Qy 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
 Db 448 GCTGGCCITATTTCATTTGGATTTTAAATACATTGAACATACTGTTCCATAGAGACGTA 507
 Qy 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThr 135
 Db 508 TGTGTGTTCTCTGACCAACCTTTAGCGGCTTACATCTATATCTATTTCTCTGCTTACA 567
 Qy 136 LysGluLeuLysAspAlaGlyValaGlyLeuLeuAlaAlaMetIleAlaValPro 155
 Db 568 AGAGAACTTTGGAACCAAGGAGGAGGACTTTTAGCTGTTTATTGCTATTGTACCA 627
 Qy 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 628 GGCTACATATCTCGGTCAGTAGCTGGATCCCTTGTATATGAAGCATGCTGTTATTGCA 687
 Qy 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195

QY 529 GlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnThrTrpAsn 548
 Db 1899 GGTATCATAGATTGCGCATGCGCAACAGGACCTCTGTGTGATACACACTCGAAC 1948
 QY 549 AenThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluValAlaTyrGlu 568
 Db 1949 AACAGCCACATCGCATGCTGCGAAAGATGCTTCCATGAACCGCGCTATATA 2008
 QY 569 IleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyr 588
 Db 2009 ATCATGAGTCCCTTGATGTCATATGTTGTGTTATTTTCGAGGAGTGATGGCTAT 2068
 QY 589 SerSerAspAlaIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThr 608
 Db 2069 TCCGGGGACGATATCAACAGTCTCTGATGCTCAGGATAGCT-----GAA 2116
 QY 609 GlyLysHis-----IleLysGluAsnAspTyrTyrThrProThrGlyGluPheArg 625
 Db 2117 GGGGAGATCCCAAGACATCGGGAGGTGACTATTTCCACCGAGGAGAGTTCGGA 2176
 QY 626 ValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyr 645
 Db 2177 GTAGACAAAGCTGGGTCTCTACTCTGTTAACTGCCTTATGTATATAAATGTCTACTAC 2236
 QY 646 ArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsn 665
 Db 2237 AGATTGGAGAAATGCAGCTAGATTTTCGACTCCCGGCTTTGACCGAACACGTAT 2296
 QY 666 AlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGluTyrThrThrGlu 685
 Db 2297 GGTGAGATTGGAATAAAGACATTAAATTCGAAGCATTTGGAGGAAGCTTTTACATCAGAG 2356
 QY 686 HisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2357 CACTGGCTTGCAGGATATATAAGTGAAGCCACTGCAACACAGG 2401

RESULT 10

ADD94789
 ID ADD94789 standard; DNA; 2417 BP.

XX AC ADD94789;

XX DT 29-JAN-2004 (first entry)

XX DE Drosophila melanogaster STT3 gene sequence.

XX KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cystic; immunosuppressive; antitense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide;
 KW fruit fly; gene; ds; STT3.

XX OS Drosophila melanogaster.

XX PN WO2003054008-A2.

XX PD 03-JUL-2003.

XX PF 18-DEC-2002; 2002WO-CA001967.

XX PR 20-DEC-2001; 2001US-00028384.

XX PA (COMP-) COMPATIGENE INC.

XX PI Perreault C, McBride K;

XX DR WPI; 2003-559122/52.

XX DR P-PSDB; ADD94790.

PT New human source of immunodominant MHC-associated peptide (SIMP) nucleic
 PT acids and proteins useful for diagnosing and treating cancers, e.g. lung
 PT or breast cancer, or for suppressing an immune response in an autoimmune
 PT disease.

XX Claim 6; SEQ ID NO 7; 66pp; English.

XX This invention relates to a novel isolated or purified human protein,
 CC termed source of immunodominant major histocompatibility complex (MHC) -
 CC associated peptide (SIMP), which is expressed ubiquitously in human
 CC cells, where the protein has the potential of generating several protein
 CC fragments binding with high affinity to a human leukocyte antigen (HLA)
 CC molecule. The invention may allow development of therapeutics with
 CC cytotoxic or immunosuppressive activity or provide sequences useful for
 CC antisense therapy or gene therapy. The source of immunodominant MHC-
 CC associated peptide (SIMP) nucleic acids, proteins and fragments are
 CC useful for diagnosing and treating cancers, for example lung cancer,
 CC intestine cancer, sarcoma, prostate cancer, testicular cancer, breast
 CC cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP
 CC proteins are also useful for modulating an immune response. Decreasing
 CC lymphoid cell proliferation is useful for suppressing an immune response
 CC responsible for an autoimmune disease or a transplant rejection. The
 CC present sequence is that of the Drosophila melanogaster STT3 gene which
 CC is related to the invention. Note: This sequence does not appear in the
 CC specification but was obtained by the indexer from GenBank.

XX SQ Sequence 2417 BP; 525 A; 628 C; 654 G; 610 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7,24e-211 Length: 2417
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 58.90% Indels: 51
 Ds: 10 Gaps: 9

US-10-028-384-12 (1-705) x ADD94789 (1-2417)

QY 16 ThrLeuLeuLysLeuLeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 Db 119 AGCCTAATCACTTCGGCATCTCTGTAATCGCTGCTGCGCGATTTCTCTCGCCCTC 178
 QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 Db 179 TTCGCGTCATCGCTTCGAGTGCATATTCATGATTTGATCCGTTCACTACACCGG 238
 QY 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 Db 239 GCCACCGCTACATGGTGCAGAAATGGTTGTACAACTTCTCAACTGCTTCGACGAGCGC 298
 QY 76 AlaTrpTyrProLeuArgGlyIleGlyGlyThrIleTyrProGlyLeuMetIleThr 95
 Db 299 GCATGGTATCCGCTCGCAGGATTTGGCGGTACCGTCTATCCCGCTGTATGATAG 358
 QY 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
 Db 359 TCCGCGGAATTCATTCGCTGCTGCACCTACTCAACATACCGGTCCATATTCGTGACATC 418
 QY 116 CysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeuThr 135
 Db 419 TCGGTGTTCTTCGGCGCGATCTTCAGTGGCTGCTACCTCCATCTCCACCTACCTGCTGACC 478
 QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
 Db 479 AAGGAGCTGTGGTCCGCGGCGCGGCTCTTCGCGCGCCAGCTTCATCGCATTCGCT 538
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 539 GGCTACATCATAGTGGTGGCTGGCTGATCTGATACATACAGGAGCATTCATTCATTCGCGC 598
 QY 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 Db 599 CTGCAGTTCACTACTTCTCTGCTGGGTGCGCTCAGTGAAGACTGGAATCCGTGTTCTGCTGCG 658

| | | | |
|----|------|--|------|
| QY | 196 | AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeu | 215 |
| DB | 659 | GCAGCAGCGCTGTCTCTACTTCTACATGGTTCGCTGGGTGGGTACTGTCTCATC | 718 |
| QY | 216 | IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg | 235 |
| DB | 719 | ATCAACCTGATACCCCTGCAGCTCTCTGACTCTCTATTATGGCAGGTACTCCCGCGT | 778 |
| QY | 236 | IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer | 255 |
| DB | 779 | CTGTGACACGCTACAGACCTCTCATCTCGGACTCTCTCTTCATCGACATCCCC | 838 |
| QY | 256 | PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly | 275 |
| DB | 839 | TTCTGTGGGATTCACACGATCAGCACCAGTGAACACATGCTGCCTGGAGTGTGTGTG | 898 |
| QY | 276 | LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe | 295 |
| DB | 899 | CTCTTATGGCGTGGCCACCTTCGCGCATTTTCAGTCTGCTCTCCGCGAACAGATTC | 958 |
| QY | 296 | GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuThrValGly | 314 |
| DB | 959 | CGGAAGCTGTTTC-----ATCGTCGGCGGATTCTGCTGGTGGCGTGGCGTTC | 1003 |
| QY | 315 | -----AlaLeuLeuMetLeuThr-----GlyLysIleSerProTyrThrGlyArgPhe | 330 |
| DB | 1004 | TTTGTGGCGCTGCTGGTGTCTCACCATCTGGGGGTGTGGCCCGGTGGAGTGCAGTTC | 1063 |
| QY | 331 | TyrSerLeuLeuAspProSerTyrAlaLysAsnIleProIleIleAlaSerValSer | 350 |
| DB | 1064 | TACTCGCTGTGGATCTGCTGCTAGCGCAAGATCCACATTCCCATCTGCATCGGTGTG | 1123 |
| QY | 351 | GluHisGlnProThrThrTyrPheAspLeuGlnLeuLeuValPheMet | 370 |
| DB | 1124 | GAGCATCAGCCACCACCTTGTCTCTCTCTTTGATCTGCATCTCTGTGTGGCC | 1183 |
| QY | 371 | PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIle | 390 |
| DB | 1184 | TTCCAGTGGAGGTGTGTACTGTCATCAGCAGATCAACGACGCGGTTCGTGGTG | 1243 |
| QY | 391 | MetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeu | 410 |
| DB | 1244 | CTGTACGCCATCAGTGGCGTTTACTTCGCTGGGTGATGTCGCTGTGATGTTCACCTC | 1303 |
| QY | 411 | AlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMet | 430 |
| DB | 1304 | ACGCCGTGTGTGCATGCTGGCGGAGTGGCCCTTTCCGGACTGTGGATGTGTCTGTG | 1363 |
| QY | 431 | LysAsnLeuAspIleSerArgPro-----AspLysLysSerLys----- | 438 |
| DB | 1364 | CAAGAGGATTCTCTAAGCGAATGGGCACAGCCATAGCGCAGCCACCGAAGTGGATGA | 1423 |
| QY | 439 | -----AspLysLysSerLys----- | 443 |
| DB | 1424 | GCTGAGGATTCCATTGAGAGAGACGCTGTACGACAGGCTGGCAACCTGAACATCGT | 1483 |
| QY | 444 | -----LysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGly | 459 |
| DB | 1484 | ACTAAGCATGANGCCACGAGGATACT-----GGCGTCAAGTCCAACTGAAGAGTATT | 1537 |
| QY | 460 | MetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrValThr | 479 |
| DB | 1538 | GTTATTTTGGCGGTCTTAATGCTGTGTATGTTCTGCTCTCCATCGCTGGGTGACC | 1597 |
| QY | 480 | SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGly---AspGlySer | 498 |
| DB | 1598 | ASCAATGCTACTTCAGTCTCTCCATGCTGTGGCTTCCCAACACAGTCAGATGGAATCC | 1657 |
| QY | 499 | ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu | 518 |
| DB | 1658 | CCCAACATTTAGACGATTCAGAGGCTTACTACTGTCTTTCCGACAACTGCCGAT | 1717 |

| | | | |
|-----------|----------|---|--|
| QY | 519 | AspAlaLysValMetSerTTPtTpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg | 538 |
| DB | 1718 | GATGCTCCGCTTATGTTCTTGGTGGGATTACCGATACCAAGATAGCGGAATGGCAACAGA | 1777 |
| QY | 539 | ThrIleLeuValAspAsnAsnThrTTPAsnAsnThrHisIleSerA-gValGlyGlnAla | 558 |
| DB | 1776 | ACGACGCTAGTGGGATATAATACGTGGAACAATAGTCACATAGCGCTGGTTGGCAAGCA | 1837 |
| QY | 559 | MetAlaSerThrGluGluLysAlaTyrGluIleMetA-gGluLeuAspValSerTyrVal | 578 |
| DB | 1838 | ATGTCCTTCAACCGGAGGAAGTCTCTACGAAATATGACATCTCTTGACGTGGACTACGTT | 1897 |
| QY | 579 | LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTTP | 598 |
| DB | 1898 | TGTGGTGAATCTTTGGCGGTGGATCGGCTATTCTGGCGATGATATCAACAAGTTCCTGTGG | 1957 |
| QY | 599 | MetValArgIleGlyGlySerThrAspThrGlyLysHis-----IleLysGluAsn | 615 |
| DB | 1958 | ATGTCGTCGAATTGCT-----GAGGCGAGAGCATCCCAAGGACATTAAGGAAAGC | 2005 |
| QY | 616 | AspTyrTyrThrProThrGlyGluPheA-gValAspA-gGluGlySerProValLeuLeu | 635 |
| DB | 2006 | GATTACTTTACGACCGCGGTGAATTCAGGGTAGATGCCGAAGGTGCTCCGGCCCTGCTC | 2065 |
| QY | 635 | AsnCysLeuMetTyrLysMetCysTyrTyrA-gPheGlyGlnValTyrThrGluAlaLys | 655 |
| DB | 2066 | AACTGCCCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTACAGA | 2125 |
| QY | 655 | ArgProProGlyPheAspA-gValA-gAsnAlaGluIleGlyAsnLysAspPheGluLeu | 675 |
| DB | 2126 | GGTCCACTCGGATATGATGCACACGTAACGCCGCTATTGGGAATAAGGACTTCGATCTG | 2185 |
| QY | 675 | AspValLeuGluGluGlyTyrThrThrGluHisTTPLeuValArgIleTyrLysValLys | 695 |
| DB | 2186 | ACCTACCTGGAGGAGCGCTACACACAGAACACTGGCTTGTTCGATCTATAGGTTGAAG | 2245 |
| RESULT 11 | | | |
| ABU | ABL13247 | ID | ABL13247 standard; cDNA; 2699 BP. |
| XX | XX | AC | ABL13247; |
| XX | XX | AC | ABL13247; |
| DT | DT | DT | 26-MAR-2002 (first entry) |
| XX | XX | XX | Drosophila melanogaster expressed polynucleotide SEQ ID NO 34223. |
| XX | XX | XX | Drosophila; developmental biology; cell signalling; insecticide; |
| KW | KW | XX | pharmaceutical; gene; ss. |
| XX | XX | OS | Drosophila melanogaster. |
| XX | XX | XX | WO2001/1042-A2. |
| XX | XX | XX | 27-SEP-2001. |
| XX | XX | XX | 23-MAR-2001; 2001WO-US009231. |
| PF | PF | XX | |
| XX | XX | XX | 23-MAR-2000; 2000US-0191637P. |
| PR | PR | XX | 11-JUL-2000; 2000US-00614150. |
| XX | XX | XX | (PERK) PE CORP NY. |
| XX | XX | XX | Venter JC, Adams M, Li PWD, Myers EW; |
| PI | PI | XX | WPI; 2001-656860/75. |
| DR | DR | XX | P-PSDB; ABB69144. |
| XX | XX | XX | New isolated nucleic acid detection reagent for detecting 1000 or more |
| PT | PT | XX | genes from Drosophila and for elucidating cell signaling and cell-cell |
| PT | PT | XX | interactions. |
| XX | XX | XX | Claim 1; SEQ ID NO 34223; 21pp + Sequence Listing; English. |

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB157737-
 CC AB172072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 2699 BP; 595 A; 695 C; 700 G; 709 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8,52e-211 Length: 2699
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 58.90% Indels: 51
 DB: 4 Gaps: 9

US-10-028-384-12 (1-705) x AB113247 (1-2699)

QY 16 ThrLeuLeuLysLeuLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 DB 128 AGCTAATACCTTGGCCATCTCTGCTAATCGCTGGTGGCCGGATTTTCCTCTCGCCTC 187
 QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAsnProTyrPheAsnTyrArg 55
 DB 188 TTCGCCGTATCATCGTTTCAGATCGATATCCATGAGTTTGATCCGCGTTCACTACCGG 247
 QY 56 ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 DB 248 GCACCGCGCTACATGTGTCAGATGTTGTGTACAACTTCTCAACTGTTTCACGAGCGC 307
 QY 76 AlaTrpTyrProLeuGlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThr 95
 DB 308 GCATGTATCCGCTCGCAGGATTTGGCGGTACCGCTCTATCCCGCTGATGATTACG 367
 QY 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
 DB 368 TCCGGCGGAATCATGGCTGTGCACGTACTCAACATACCGCTTCATATCGTGACATC 427
 QY 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThr 135
 DB 428 TGGCTGTTCTGGCGCGGATCTTCAGTGCCTGACCTCCATCTCCACCTACCTGTGACC 487
 QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValPro 155
 DB 488 AAGAGCTGTGTCGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCTGCTGCT 547
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 DB 548 GGTACATCAGTAGGTCGTGTGATCGTACGATACGATACGAGGGGATTCGCAATATCGCC 607
 QY 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 DB 608 CTGCAGTTCACTTCTGTTGGTGGCTCAGTGAACATGATCGTATCGTGTCTGGTCG 667
 QY 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerTrpGlyGlyTyrValPheLeu 215
 DB 668 GCGCGAGCGCGCTTGTCTTCTTCTACATGTTGTGCTGCTGCTGCTGCTGCTGCTGCT 727
 QY 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 DB 728 ATCAACCTGATACCTGTCACGCTCTTGTACTGTCTCATATGCGGAGGTACTCGCGCGT 787
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 DB 788 CTGCTGACCGAGCTACAGCACCTTCTACATCTCTGGGACTGCTGTTCTCCATGCAGATCCC 847
 QY 256 PheValGlyPheGluProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275

DB 848 TTCGGGGATTCACACCGCATCGCACCAAGTGAACACATGGCTGGCTGGAGTGTGTGTG 907
 QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPhe 295
 DB 908 CTCTTATGGCGGTGGCCACCTTGGCCATTTGAGTTCGGTCTGGCGCAACGAGTTC 967
 QY 296 GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuLeuValGly--- 314
 DB 968 CGGAAGCTGTTC-----ATCGTCGCGGATTTGCTGGTGGCGTGGCGTC 1012
 QY 315 -----AlaLeuLeuMetLeuThr-----GlyLysIleSerProTyrThrGlyArgPhe 330
 DB 1013 TTTGTGGCGGTGGTGTCTCACCATGCTGGCGGTGTGGCCCGCTGGAGTGGAGCTTC 1072
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSer 350
 DB 1073 TACTCGCTGTGGGATCTAGTGGCTAGCCCAAGATCCACATCCCATTCATTGCTGCTGTCG 1132
 QY 351 GluHisGlnProThrThrTrpSerSerTyrPheAspLeuGlnLeuLeuValPheMet 370
 DB 1133 GAGCATCAGCCACCATTTGGTCTCTGTTCTTCTTGTGATCTGCACATCTCTGGTGGCGC 1192
 QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 390
 DB 1193 TTCCAGTGGGAGTGTGTACTGCATCAAGCAGATCAACGACGAGCGGTTTCTGCTGCTG 1252
 QY 391 MetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeu 410
 DB 1253 CTGTACGCCATCAGTGGGTTTACTTCGCTGGTGTGATGGTGGCTTTGATGTTGACCTC 1312
 QY 411 AlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMet 430
 DB 1313 ACGCCGTGGTGTGCATGCTGGCGGAGTGGCTTTTTCGGGACTGTTGGATGTTCTCTG 1372
 QY 431 LysAsnLeuAspIleSerArgPro----- 438
 DB 1373 CAAGAGGATTCGTAAAGCAATGGGCACACCCATAAGCGACCCAGCCAGTGGATGAA 1432
 QY 439 -----AspLysLysSerLys----- 443
 DB 1433 GCTGAGGATTCATTGAGAGAGAGACGCTGTACGACAAAGCTGGCAAGCTGAAGCATCGT 1492
 QY 444 -----LysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGly 459
 DB 1493 ACTAAGCATGATGCCACGAGGATACT-----GGCTGAGCTCCCAACCTGAAGAGTATT 1546
 QY 460 MetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThr 479
 DB 1547 GTTATTTGGCCGTCTTAATGCTGTTGATGATGTTGCTGCTCCATGCTGACGTTGGTACC 1606
 QY 480 SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGly---AspGlySer 498
 DB 1607 AGCAATGCTACTCCAGTCCCTCCATGTTCTGGCTTTCACCAACAGTCAAGATGGATCC 1666
 QY 499 ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
 DB 1667 CGCAACATTTAGACGATTTTCAGAGGCTTACTACTGGCTTTTCGAGAACACATGCCGAT 1726
 QY 519 AspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
 DB 1727 GATGCTCGCTTATGTTCTTGTGGGATTCGATACGATACGATGCGGGAATGGCAACAGA 1786
 QY 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
 DB 1787 ACCACGCTAGTGGATTAATACGTGGAACAATPAGTCACATAGCGTGGTGGCAAGGCA 1846
 QY 559 MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal 578
 DB 1847 ATGCTTCAACCGGAGGAGTCTCTACGAAATTTATGACATCTCTTACGCTGGACTACGTT 1906
 QY 579 LeuValIlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrp 598
 DB 1907 TTGGTATCTTTGGCGGTGTGTCGCTATTCTTGGCGATGATATCAACAGTTCCTGCTGG 1966

QY 599 MetValArgIleGlySerThrAspThrGlyLysHis-----IleLysGluAen 615
 Db 1967 ATGGTCCGAATTGCT-----GAGGAGAGCATCCCAAGGACATTAAAGGAAGC 2014
 QY 616 AspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeu 635
 Db 2015 GATTACTTTACCGACCGCGGTGAATTACGGGTAGATCCGAAGTGTCTCGGCGCTGCTC 2074
 QY 636 AsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLys 655
 Db 2075 AACTGCCCTTATACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTACAGA 2134
 QY 656 ArgProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeu 675
 Db 2135 GGTCCATCTGGATATGATCGCACAGTAACCGCGTCAATGGGAATAAGGACTTCGATCTG 2194
 QY 676 AspValLeuGluGluGlyTyrThrGluHisTrpLeuValArgIleTyrLysValLys 695
 Db 2195 ACCTACCTGGAGAGCGCTACACACAGAACTGGCTGTTCGCATCTATAGGTGAAG 2254
 RESULT 12
 ID ABL13246/c
 XX ABL13246; standard, cDNA; 4922 BP.
 AC ABL13246;
 XX 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34220.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 XX
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR P-PSDB; ABB69143.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 34220; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4922 BP; 1426 A; 1210 C; 1142 G; 1144 T; 0 U; 0 Other;
 Alignment Scores: 7.9e-203 Length: 4922
 Pred. No.:

Score: 2103.00 Matches: 415
 Percent Similarity: 67.97% Conservative: 113
 Best Local Similarity: 53.34% Mismatches: 140
 Query Match: 56.88% Indels: 112
 DB: 4 Gaps: 11
 US-10-028-384-12 (1-705) x ABL13246 (1-4922)
 QY 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
 Db 3795 AGCCTAAATCACCTTCGCCCATCTCTGTAATCCCTCGGCGCGGATTTCTCTCGCTC 3736
 QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 Db 3735 TTCGCGCTCATCCGTTTCGAGTCGATTATCCATGAGTTGATCGGTGGTTCACTACCGG 3676
 QY 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 Db 3675 GCCACCGCTACATGTCAGAAATGGTTGGTACAACTTCCTCACTGGTTCCAGAGCGC 3616
 QY 76 AlaTyrProLeuGlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThr 95
 Db 3615 GCATGGTATCCCTCGGCGAGGATTGTGGCGGTACCGTCTATCCGCGCTCATGATTACG 3556
 QY 96 SerAlaAlaIleTyrHisValLeuHisPheHisIleThrIleAspIleArgAsnVal 115
 Db 3555 TCCGCGGAATCCATTGGCTGCTGCACGTACTCAACATACCGGTCCATATTCGTGACATC 3496
 QY 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThr 135
 Db 3495 TCGGTGTTCTCGGCCCGCATCTTCAGTGGCTGACCTCCATCTCCACCTACCTGCTGACC 3436
 QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
 Db 3435 AAGGAGCTGTGTCGCGCGCGCGCTCTTCGCGCGCAGCTTCATCGCATCGTGCCT 3376
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 3375 GGTATACATCAGTAGGTGCGTGGCTGATCGTACGATACGAGGCAATGCCATATTCGCC 3316
 QY 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 Db 3315 CTGCAGTTACACCTACTTCTCTGGTGGTGGCTCAGTGAAGACTGGATCGCTTCTGTGTCG 3256
 QY 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeu 215
 Db 3255 GCGCAGCGCGCTTGTCTTACTTCTACATGGTGTGCGCGTGGGGTGGCTAGTGTTCATC 3196
 QY 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 Db 3195 ATCAACCTGATACCCCTGACCGCTCTTCGTACTGCTCATTATGGGCAGGTACTCGCGCGT 3136
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 Db 3135 CTCTCAGCAGGTACAGCACCTTCTACATCTCTGGGACTGCTGTCTTCATGAGATCCCC 3076
 QY 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 Db 3075 TTCGTGGGATTCACACGATACGACCATGAGTGAACATGGCTGCGCTGGAGTGTTCGTG 3016
 QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnIlePhe 295
 Db 3015 CTCCTTATGGCGGTGGCCACCTTGGCCCATTTGGCATGCTCGTCTCGCGCAACAGATTC 2956
 QY 296 GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuThrValGly--- 314
 Db 2955 CGGAAGCTGTTC-----ATCTGCGCGGAGTTCTGTGGCGCTTGGCGTTC 2911
 QY 315 -----AlaLeuLeuMetLeuThr-----GlyLysIleSerProTyrThrGlyArgPhe 330
 Db 2910 TTTGTGCGCGTGTGTTGCTCACCATGCTGGCGGTGTTGGCGCGGTGGAGTGGACGCTTC 2851
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSer 350

compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of C. albicans cells and for treating infection by C. albicans. The present sequence is that of an essential Candida albicans gene used in the method of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to perwent by the European Patent Office

[illegible]

| | | |
|------------------------|-----------|-------------------|
| Alignment Scores: | | |
| Pred. No.: | 5.06e-200 | Length: 2356 |
| Score: | 2070.50 | Matches: 353 |
| Percent Similarity: | 72.33% | Conservative: 122 |
| Best Local Similarity: | 75.20% | Mismatches: 174 |
| Query Match: | 56.00% | Indels: 23 |
| DR: | 6 | Gaps: 8 |

US-10-028-384-12 (1-705) X ABZ32036 (1-2256)

| | | | |
|----|-----|---|-----|
| Qy | 7 | LeuArgLeuSerTyrGluLysGlnAspThrIleuLeuLysLeuLeuIleLeuSerMetAla | 26 |
| Db | 40 | TTGGGGATAGATGTTGAACCTGATTAGAGTTTATAAAGTTTATATATTTATATCGATA | 99 |
| Qy | 27 | AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis | 46 |
| Db | 100 | GCAGGTGCAGCTATTTCTTCGTTTATTTCCGTGATTCGATTTGAAAGTAGTATTATTCAT | 159 |
| Qy | 47 | GluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluLysPheTyr | 66 |
| Db | 160 | GAATTCGATCCCTGGTTCAAATTCGAGCAACCAAAATATTAGTCACTCATTCCTTTAT | 219 |
| Qy | 67 | LysPheHisAsnTrpPheAspArgAlaTrpTyrProLeuGlyArgIleLeuGly | 86 |
| Db | 220 | GAATTTTGAATTTGGTTTCATGATAGAACTTGGTACCCATTGGGAAGAGTCACTGGTGT | 279 |
| Qy | 87 | ThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHis--- | 105 |
| Db | 280 | ACTTTATATCCCGGTTAATGGTGACTTTTCAGGTGCCAATTTGGCATATTTACGTGATGT | 339 |
| Qy | 106 | PheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerSer | 125 |
| Db | 340 | TTTGCCTTACCCGTTGATATTAGAAATATTGTGTTTATTAGCACCCAGTTTCTCGGA | 399 |
| Qy | 126 | PheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyAlaGlyIleu | 145 |
| Db | 400 | TTAACTGCATTTTGACTTATTTTGTGATAAGAAATGAAGGATCTAGTGCAGGATTA | 459 |
| Qy | 146 | LeuAlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAlaGlySer | 165 |
| Db | 460 | TTGGCAGCTATATTATGGGATTCGCCCGGTTATTTTCAAGATCAGTGGCTGTCT | 519 |
| Qy | 166 | TyrAspAsnGluGlyIleAlaIlePheCysMetIleLeuThrTyrMetTrpIleLys | 185 |
| Db | 520 | TATGATTAATGAAGCAATTCGCCATTACTTTATTAATGGCAACATTTATTTCTGGATAA | 579 |
| Qy | 186 | AlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMet | 205 |
| Db | 580 | TCAATGAANAATGGTTTCAGTTTCATGCCCAATTCACAGCATTTCTATTTCTATATG | 639 |
| Qy | 206 | ValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIleProLeuHisValLeuVal | 225 |
| Db | 640 | GTTAGTGCTTTGGGTGGATATGTTTTCATTACCAATTTGATTCCATTACACGATTATTGC | 699 |
| Qy | 226 | LeuMetLeuThrGlyArgPhePheSerHisArgIleTyrValAlaTyrCysThrValTyrCys | 245 |
| Db | 700 | TTGATTTTCATGGTTCGTTAATATGCAAACTTTACACTGCTTATACATGTATGCC | 759 |
| Qy | 246 | LeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerSer | 265 |
| Db | 760 | TTGGGTACTTTGGGCATCAATGTCAGATTCCAATTCGTTGGGTTTTTACCAAAGATCAAA | 819 |

QY 611 HisIleYsGluAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGly 630
 DB 1891 GAAATCAAGAAAGAGACTACTTACTACCGAGGAGATATAAGTGGATGAAGTCCA 1950
 QY 631 SerProValLeuLeuAsnCysLeuMetTyrIleYsMetCysTyrTyrArgPheGlyGlnVal 650
 DB 1951 TCATCGCAATCAAGAATCTCTTGATGATATAAGTTATCGTATCATAGATTCACTGAATTG 2010
 QY 651 TyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGluIle---Gly 669
 DB 2011 TTTGGAGGTAGA-----GATGGTGTGATAGATTGAAACCAACAAATCCAGGCC 2061
 QY 670 AsnLysAspPheGluLeuAspValLeuGluGlyTyrThrGluHisTyrPheLeuVal 689
 DB 2062 AATGAAGTACCGAAATGAATGTTGTTGAAGAAGCCCTTCATCATCAGAAATTCGATTGTG 2121
 QY 690 ArgIleTyrIleValLysAspLeuAspAsnArgGly 701
 DB 2122 AGAATTACAAAGTTAAGATTGGATATGTTGGT 2157

RESULT 14

ID ABT20789 standard; DNA; 2232 BP.
 XX AC ABT20789;

XX 16-APR-2003 (first entry)

XX Aspergillus fumigatus essential gene #3147.

XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 KW cancer; contamination; biofilm; antibody; immune response; ds.

XX Aspergillus fumigatus.

XX WO200286090-A2.

XX 31-OCT-2002.

XX 23-APR-2002; 2002WO-US013142.

XX 23-APR-2001; 2001US-0285697P.

XX 27-APR-2001; 2001US-0287066P.

XX 05-JUN-2001; 2001US-0295890P.

XX 09-JUL-2001; 2001US-0303899P.

XX 31-AUG-2001; 2001US-0316362P.

XX (ELIT-) ELITRA PHARM INC.

XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
 XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of Aspergillus
 PT fumigatus, useful for treating or preventing infections by A. fumigatus,
 PT or for treating a non-infectious disease in a subject e.g. cancer.

XX Disclosure; Page; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and

CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention

XX Sequence 2232 BP; 473 A; 605 C; 545 G; 609 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.15e-194 Length: 2232
 Score: 2010.00 Matches: 374
 Percent Similarity: 70.62% Conservative: 126
 Best Local Similarity: 52.82% Mismatches: 170
 Query Match: 54.37% Indels: 38
 DB: 8 Gaps: 8

US-10-028-384-12 (1-705) x ABT20789 (1-2232)

QY 17 LeuLeuLysLeuLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPhe 36
 DB 61 CTTCTGGCGATTATTATCTCTGTACCATTCAGCAGCAGCAGTTCGAGTCGATCTTTC 120
 QY 37 AlavalLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThr 56
 DB 121 AGCGTTATCGGCTTCGAGAGTATCATCCAGAAATTCACCGCGTGTCAACTCCGAGCA 180
 QY 57 ThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAla 76
 DB 181 ACAAAATACTTAGTACAGATGGTTCTATAGCTTTGGGATGTTGTATGACGGGAACA 240
 QY 77 TrpTyrProLeuGlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSer 96
 DB 241 TGGCATCTCTTGGGACGCTGTCCCGGTGGCAGCTATATCCCGTCTCATGGTCACGAGC 300
 QY 97 AlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCys 116
 DB 301 GCGGTGATCTACCATATCTTTCGATTCTTACTATCCCGCTCGATATTCGCAACATCTGC 360
 QY 117 ValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLys 136
 DB 361 GTCTACTGGCCCGCAGGATCTCCGGCGCTGACTGCATGGCAATGTACTTGTGACATCC 420
 QY 137 GluLeuLysAspAla---GlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
 DB 421 GAGATGCTCTCTCGCCATCTGCAGGTCTTCTTGCAGCAGCTTTCATGGGAATCGCCCT 480
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 DB 481 GGTTACATCTCCCATCAGTTGCTGGAAGTACGATACGAGGATTCGCACTTTCTTG 540
 QY 176 MetLeuLeuThrTyrTyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrPala 195
 DB 541 CTTGTGTTCATCTTCTTCTATGGATCAAGGCTGTCAAAATGGGTCTATCATGTGGGA 600
 QY 196 AlalysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeu 215
 DB 601 GCGGTGACCGCACTATTCTACGCTACATGGTTCGCGATGGGGTGGGTATGCTTCATT 660
 QY 216 IleAsnLeuLeuProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 DB 661 ACGAACCTGATCCCGCTGCAGCTTTTGTCTTCTGTGATGGGTAGATACAGCACTCCG 720
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 DB 721 ATTACATTAGTATACCATGATGGTATGGCGCTGGGACTTTTGGTAGCATGCAGATCC 780

QY 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 Db 781 TTGGTGGGATTTTGGCTTATCCGAAACAGCAGCAGCAGCAGCTCCGGCTTGGTGTCTTCGGC 840
 QY 276 LeuGlyGlnHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
 Db 841 CTGCTTCAGCTGTGGCTTCGGCAGTTGTCCGACAGTTCTCCACAGCAGCTTC 900
 QY 296 GluValLeuPheArgSerValile-----SerLeuValGlyPheVal 309
 Db 901 CAGACACTTTCAGCGCCATGCTTTCATCACCCTTCGCTCGCTTCGCTTGGCTAGTT 960
 QY 310 LeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysLysSerProThrPheGlyArg 329
 Db 961 GTTCTGACTGTG-----ACGGAGGTGATCGCTCTCTGAGCGGCGCA 1002
 QY 330 PheTyrSerLeuLeuAspProSerTyrAlaLysAsnAenileProilelleAlaSerVal 349
 Db 1003 TTCTACTCTTGTGGGACTGGCTATGCCAAATCCACATCCCATCATTCCTCAGTC 1062
 QY 350 SerGluHisGlnProThrThrTrpSerSerTyrTyrPheAspLeuGlnLeuValPhe 369
 Db 1063 TCGGAACACACAGCCAGCTTGGCCAGCTTCTTTCGATCTGAACCTTCCTCATCTGG 1122
 QY 370 MetPheProValGlyLeuTyrCysPheSerAsnLeuSerAspAlaArgilePheile 389
 Db 1123 CTTTTCCTCCGGCAGGTGTCTACATGCTTCTCGTACCTCAAGCAGCAGCATGTCTTCGTC 1182
 QY 390 IleMetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuVal 409
 Db 1183 ATTATCTACTCGTCTTGGGACTTCTCGCGGTGTATGGTCCGACTAATGCTGACC 1242
 QY 410 LeuAlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyr 429
 Db 1243 TTGACCCCTATTGTGTGTCTCCGCTGTCTGGCGCTGTCTCCATCTCGACACCTAT 1302
 QY 430 MetLysAsnLeuAspIleSerArgProAspLysLysSerLysGlnGlnAspSer--- 448
 Db 1303 ATGGCGACT--ACCTCCCGACACACCGCTCTGAAGCGGAAACCGAATGAAGACTCGTCT 1359
 QY 449 -----ThrTyrProIleLys 453
 Db 1360 TCACAACTCTTCGCTCAGTTCGGAAGCCCAATGTGAATCACCTCCATGTTTCTTAAG 1419
 QY 454 IleGluValAlaSerGlyMetIleLeuValMetAlaPhePheIleThrTyrThrPhe 473
 Db 1420 ATTATAGTACGGCTGTCTGTCTGCTC-----TACCTGCTCTGTGTTCGCG 1467
 QY 474 HisSerThrTrpValThrSerGluAlaTyrSerProSerIleValLeuSerAlaArg 493
 Db 1468 CACTGCACCTGGGTATACATCAATGATGATCTCTCTTCGCTTCTGCTGGTTCGT 1527
 QY 494 GlyGlyAspGlySerArgilellePheAspAspPheAspGluAlaTyrTyrTrpLeuArg 513
 Db 1528 ATGCTCTGACGAAGCCAAATACATCATGACGATATATCGTACGCTTACTCTGGTTCGT 1587
 QY 514 HisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThr 533
 Db 1588 CAGAATACTCTCAGAAACGCCAAATCATGTATGTTGGGATATGGGTATCAAAATCGGT 1647
 QY 534 AlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSer 553
 Db 1648 GGCATGCGGACCGCCCACTTGGTGTGACACACACCTGGAACACACCATATTGCT 1707
 QY 554 ArgValGlyGlnAlaMetAlaSerThrGluGlyLysAlaTyrGluIleMetArgGluLeu 573
 Db 1708 ACGTTTGGTAAAGCGGATGAGCTCAGCGAGGAAGTCAAGTACCCCATCTCCGCCAGCAT 1767
 QY 574 AspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIle 593
 Db 1768 GATGTGCATTACGTGTGGTGGGTCTCGTGGTGGTCTCTAGGTATTCGGCAGTACATT 1827
 QY 594 AsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLys 613

Db 1828 AACAAATTTCTATGATGGTCCGTATCGCCGAAGGTATC---TGGCCCGATGAGTTAAA 1884
 QY 614 GluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProVal 633
 Db 1885 GAGCGGGACTTCTTACTTCGACGGGTGAATATCGTTCGACGATGAGGACGCCCAACT 1944
 QY 634 LeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGlu 653
 Db 1945 ATGCGCAACAGCTTGTGATATAAATGCTTATTCACTTCACTCTCTCTCCCGTCG 2004
 QY 654 AlaLysArgProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPhe 673
 Db 2005 GCCCAA-----GCTGTGCGACCGCTCGGTGGTCAAAACTTCCACAGAGGCCCT 2055
 QY 674 GluLeuAspValLeuGluGluGlyTyrThrThrGluHisTrpLeuValAlaGlyLeuTyrLys 693
 Db 2056 CAGCTCTCTACACTCGAAGAGCTTTCACGAGCGAGAACTGATCATCTGTTATCTACAG 2115
 QY 694 ValLysAspLeuAspAsnArgGly 701
 Db 2115 GTCAAGGATCTTGACACCTTGGC 2139
 RESULT 15
 AD36900
 ID ADB69900 standard; DNA; 2466 BP.
 XX
 AC ADB69900;
 XX
 DT 04-DEC-2003 (first entry)
 DE
 XX
 DE C. neoformans open reading frame SEQ ID NO:2305.
 XX
 KW ds; gene; fungicide; gene therapy; infection.
 XX
 OS Cryptococcus neoformans.
 XX
 FN WO2003052076-A2.
 XX
 PD 26-JUN-2003.
 PF 17-DEC-2002; 2002WO-US040225.
 XX
 PR 17-DEC-2001; 2001US-0341261P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Zamudio C, Eroshkin AM;
 XX
 XX WPI; 2003-533017/50.
 DR P-FSDB; ADB70261.
 XX
 XX New nucleic acid, useful for preparing a composition for treating an infection caused by Cryptococcus neoformans.
 XX
 PS Claim 2; SEQ ID NO 2305; 136pp; English.
 XX
 CC The invention relates to a novel purified or isolated Cryptococcus neoformans nucleic acid molecule comprising a sequence encoding a polypeptide comprising a sequence not given in the specification. A polynucleotide of the invention has fungicide activity, and may have a use in gene therapy. The nucleic acid is useful for preparing a composition for treating an infection caused by Cryptococcus neoformans.
 CC The present sequence represents a C. neoformans sequence of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 2466 BP; 508 A; 678 C; 580 G; 700 T; 0 U; 0 Other;
 Alignment Scores: 1.15e-185 Length: 2466
 Pred. No.: 1930.00 Matches: 373
 Score:

us-10-028-384-12.rng

Thu Dec 16 16:25:00 2004

QY 671 sAspPheGluLeuaspValLeuGluGluGlyTyrThrThrGluHisTrpLeuValArgIl 691
 Db 2271 CAGTGTACTCTTGATACCTTTGACGAAGCGTTCCACATCCGAAATTTGGATCGTCAGGAT 2330
 QY 691 eTyrlYsVallylsAspLeuaspAsnArgGly 701
 Db 2331 CTACAAGGTCAGAAGGAGATCCCATTTGA 2361

Search completed: December 14, 2004, 20:07:11
 Job time : 905.511 secs

QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 DB 100 TTTGCAGTCATCAAAATTTGAGTCTATATCCATGAATTCGACCCCTGGTTCAATATAGG 159
 QY 56 ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArg 75
 DB 160 GCTACCAAAATATCTCGTCAACAATTCGTTTACAAAGTTTGAACGTGGTTGACACCGT 219
 QY 76 AlaTrpTyrProLeuGluArgIleIleGlyGlyThrIleTyrProGluMetIleThr 95
 DB 220 ACTTGGTACCCCTCGGAAGGGTTACTGAGGAGCTTTATATCTCGGTTGATGACGACT 279
 QY 96 SerAlaAlaIleTyrHisValLeu--HisPhePheHisIleThrIleAspIleArgAsn 114
 DB 280 AGTCGGTTCATCTGGCAGCCCTCGCACTGGTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 339
 QY 115 ValCysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeu 134
 DB 340 GTTTGTGTGCTATTTGGCCCACTATTTCTGGGTCACCCCTGGGCACTTACGAATTT 399
 QY 135 ThrLysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValVal 154
 DB 400 ACGAAACAGATTAAAGATGCGCAGCGCTGGCTTTGGCTGCTGCTTTATAGCCATTGTC 459
 QY 155 ProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePhe 174
 DB 460 CCCGTTATATATCTAGATCAGTGGCGGGTCTCTAGATANTAGGCCATTGCCATTACA 519
 QY 175 CysMetLeuLeuThrTyrMetTrpIleLysAlaValIleThrGlySerIleCysTrp 194
 DB 520 CTATTAAATGTCATTTTCATGTTTGGATTAAAGCCCAAAAGACTGGCTCTATCATGCA 579
 QY 195 AlaAlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPhe 214
 DB 580 GCAACGTGTGCGAGCTTTATCTACTTCTACATGGTGTGCGCTGGGTTGATCGTGTTC 639
 QY 215 LeuIleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHis 234
 DB 640 ATCAACCACTGATCCCACTCCATGCTCTTTTCTGATTTGATGGCAGATATTCGTCC 699
 QY 235 ArgIleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIle 254
 DB 700 AAACGTGATCTGCTACACCATCTGTCAGCTATTGGAACTGTTCATCCATCGAGATC 759
 QY 255 SerPheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPhe 274
 DB 760 CCATTTGTGGT--TCCTACCTATCAGGTCTTAACGACCATATGGCCGCTATGGGTGTTTC 819
 QY 275 GlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGln 294
 DB 820 GGTGTGATTCAGATTGCGCTTCTGGTGACTCTCGTAAGGGCCAAATCAGCACACTAAG 879
 QY 295 PheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGly 314
 DB 880 TTTAAAGTCATCATG-----ATGGTTTCTCTG-----TTTTGATCTTGGTCTCGT 927
 QY 315 AlaLeu-----LeuMetLeuThrGlyLysIleSerProTrpThrGlyArgPhe 330
 DB 928 GTGGTCGAGCTTCTGCTTGACTATATGGGTTGATTTGCCCTTGGACTGGTAGATTT 987
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAlaIleProIleIleAlaSerValSer 350
 DB 988 TATTCTGTTATGGATACCACTACGCAAAATCCACATTCCTATCATTTGCTCCCTTCC 1047
 QY 351 GluHisGlnProThrThrTrpSerTyrPheAspLeuGlnLeuLeuValPheMet 370
 DB 1048 GAAATCAACCCGTTGCGGGCGCTTCTTCTTTGATACCCACTTTTGTATCGGCTA 1107
 QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 390
 DB 1108 TTCCCCCGCGGTGATTCTTACTATTTCTCGACTTGAAGAGCAGCAGCTTTTGTCTATC 1167
 QY 391 MetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeu 410

DB 1168 GCTTACTCGTTCGTGTTGCTACTTTGCGGTGTTAGTGTAGATGATGTTGACTTTC 1227
 QY 411 AlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMet 430
 DB 1228 ACACCAAGTCATCTGTGTCGCGCGCTGCGCATTTGTCCAAGATATTTGACATCTAC--- 1284
 QY 431 LysAsnLeuAspIleSerArgProAspLysLysSerLysLysGlnGlnAspSerThrTyr 450
 DB 1285 -----CTGGATTTCAACACAGTGCAGCAAA-----TAC 1314
 QY 451 ProIleLysIleGluValAlaSerGlyMetIleLeuVal-----MetAlaPhePhe 467
 DB 1315 GCCATCAAAACCTGGCGCACTACTGCGCAAAATTTGCTTTCCGGATCATTCATCTTTAT 1374
 QY 468 LeuIleThrTyrThrPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSer 487
 DB 1375 TTGTATCTTTTCGTTCTTCCATTTCTACTTTGGTTACAGAACTGCATACCTCTCTCTCT 1434
 QY 488 IleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspPheArgGlu 507
 DB 1435 GTTGTGTTTGGCATCACAAACCCAGATGGTAAATTTGGCGTTGATCGACGACTTTCAGGAA 1494
 QY 508 AlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAsp 527
 DB 1495 GGTACTATTGGTTAAGATGAATCTCTGATGAGGACAGTAAGTTGCAGCGTGTGGAT 1554
 QY 528 TyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrp 547
 DB 1555 TAGCGTTACCAAAATTTGGTGGCATCGCAGACAGAACCACTTTAGTCGATACACACAGTGG 1614
 QY 548 AsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyr 567
 DB 1615 AACATATCTACATCGCCATCGTTGGTTAAAGCCATGCTTCCCTGAAGAGAAATCTTAC 1674
 QY 569 GluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGly 587
 DB 1675 GAAATTTCTAAAGACATGATGTCGATATTGTCTTGTGTCATCTTGTGTCCTAATTGGG 1734
 QY 588 TyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAsp 607
 DB 1735 TTTGTGTTGTTGATGATCAACAAATTTCTTGGATGATCAGATTAGCGAGGGAATC--- 1791
 QY 608 ThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAsp 627
 DB 1792 TGCCCAAGAGAGATAAAGAGCGTGTATTTCTATACCCAGAGGAGAAATACAGATAGAT 1851
 QY 628 ArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPhe 647
 DB 1852 GCAAGGGCTTCTGAGACCATGAGGACTCGCTACTTTTACAGATGCTCTACAAAGATTTC 1911
 QY 648 GlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGlu 667
 DB 1912 CCACAAATTTCAATGTTGGTGGCA-----GCCACTGACAGAGTGGCTCAACAAATG 1962
 QY 668 IleGlyAsnLysAspPhe---GluLeuAspValLeuGluGlyTyrThrThrGluHis 686
 DB 1963 ATCACACCATTAGACGTCCACCATTAGACTACTTTCGACGAAGTTTACTTCGGAAGAAC 2022
 QY 687 TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArgGly 701
 DB 2023 TGGATGGTTAGAAATATATCAATTGAAGAAGATGATGCCCAAGT 2067

RESULT 2

US-09-270-767-11648
 : Sequence 11648, Application US/09270767
 : Patent No. 6703491
 : GENERAL INFORMATION:
 : APPLICANT: Homburger et al.
 : TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
 : FILE REFERENCE: File Reference: 7326-094
 : CURRENT APPLICATION NUMBER: US/09/270.767
 : CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 11648
 ; LENGTH: 1386
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-11648

Alignment Scores:

Pred. No.: 4,94e-148 Length: 1386
 Score: 1298.00 Matches: 239
 Percent Similarity: 88.54% Conservative: 39
 Best Local Similarity: 76.11% Mismatches: 36
 Query Match: 35.11% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-270-767-11648 (1-1386)

QY 9 LeuSerTyrGluLysGlnAspThrLeuLeuLysLeuLeuLeuLeuSerMetAlaVal 28
 DB 444 CTCACCTGGGACAAACAGGACGCTGGTCAAGCTGGCCATTCTCATCTGGGAGCGGTT 503
 QY 29 LeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValHisGluPhe 48
 DB 504 TTATCAATTGGCACAGCGCTTCTCTGTGCTGGATTCGAAAGCGTATCCATGAGTTC 563
 QY 49 AspProTyrPheAsnTyrArgThrArgPheLeuAlaGluGluGlyPheTyrLysPhe 68
 DB 564 GATCGGTACTCACTACCGCACCGCGGTTCTTGGCGGAGGAGGCTTTTACAAGTTC 623
 QY 69 HisAsnThrPheAspArgAlaTyrPheLeuGlyArgLeuGlyThrLeu 88
 DB 624 CACAACCTGGTTCGATGACCGGCGCTGTGTATCCCTGGCGCGCATCATCGCGCACATC 683
 QY 89 TyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHisPheHisIle 108
 DB 684 TATCCGGGCTGATGCTCACCTCGCGGCGCTGTACCGCTGTGTGGTGTCTCAATGTG 743
 QY 109 ThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPheThrSer 128
 DB 744 ACCATCGACATACGGAACGCTGGGTCTCTCTGGCGCGCTTCTCTCTCTGTGACGACG 803
 QY 129 IleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyAlaGlyLeuLeuAla 148
 DB 804 CTGGTGACTACGCGCTTCAACAGGAGATACACAGCACTGGAGCTGGATGGTGGCGGC 863
 QY 149 AlaMetIleAlaValProGlyTyrIleSerArgSerValAlaGlySerTyrAspAsn 168
 DB 864 GCTTTGATATCATCGTTCCCGGTATATCTCTCGATCCGTGGCGGATCGTACGACAAT 923
 QY 169 GluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTyrIleLysAlaValLys 188
 DB 924 GAAGGCATCGCATTTCTGATGCTCTTCACTACTATTGTGGATCAAGCGGTAAAG 983
 QY 189 ThrGlySerIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMetValSer 208
 DB 984 ACGGACAGCATCTTTGTGCGGTATGTGCGCATTTGGCCTACTTCTATATGTCTCTCTCG 1043
 QY 209 TrpGlyGlyTyrValPheLeuLeuAsnLeuLeuProLeuHisValLeuValLeuMetLeu 228
 DB 1044 TGGGTGGCTATGCTTCTCTGATTAACCTAATCCCGCTGCACGTGTGGCGCTGATGATC 1103
 QY 229 ThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeuGlyThr 248
 DB 1104 ACCGACGCTTTCTCGCAGAGATCTATACGATACACAGCAGCTATATCTGCTCGGCACC 1163
 QY 249 IleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerGluHisMet 268
 DB 1164 ATTCTGTGATGCAGATCTCTTTGTGGATTCACACCATCCAGAGCTCCGACACATG 1223
 QY 269 AlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSer 288
 DB 1224 CTGGCACTGGGAACCTTTGGGCTGTGCCAGATTCACGCTTTCGTGCACTATCTCGGCTCG 1283

QY 289 LysLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPhe 308
 DB 1284 CGCATTCCTCCAGGATCATCTCGATCTGCTTCAAGACGTTGGTTCCAGTGTGTTGACT 1343
 QY 309 ValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLys 322
 DB 1344 GTGTGTTGCTGCTGGGTACCTGCTCACGCTTACCGGAAA 1385

RESULT 3

US-09-270-767-12856
 ; Sequence 12856, Application US/09270767
 ; Patent No. 6703491

GENERAL INFORMATION:

; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 12856
 ; LENGTH: 914
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-12856

Alignment Scores:

Pred. No.: 1,14e-107 Length: 914
 Score: 965.00 Matches: 175
 Percent Similarity: 92.04% Conservative: 10
 Best Local Similarity: 87.06% Mismatches: 16
 Query Match: 26.10% Indels: 0
 DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-270-767-12856 (1-914)

QY 502 PheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLys 521
 DB 2 TTCATGACTTCCGCGAGGCTACTACTGGCTGCAGATACCGCATGGCCATCGGAGACGCTCG 61
 QY 522 ValMetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeu 541
 DB 62 ATATGTCCTGGTGGGACTACGCTACCATGATACCGCATGGCCATCGGACGATATTA 121
 QY 542 ValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSer 561
 DB 122 GTGATTAACAATCTTGGAAACAACACATATATCCGCGTGGCCAGCGGATGGCTCT 181
 QY 562 ThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIle 581
 DB 182 TCGGAGGAGAAAGCCTACGAGATATAGGGAACTGGATGTGGACTACCTTCTCGTGTAT 241
 QY 582 PheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArg 601
 DB 242 TTCGAGGGCTCACTGCTACTCATCGACGATATCAACAAGTTCCTGTGGATGGTGGC 301
 QY 602 IleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThr 621
 DB 302 ATTGGCGGACGACGATCTGTGGCGCACATCCGCGAAAGGACTACTATGGGCGCAAC 361
 QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
 DB 362 GGAAGTTCGAGTGGCAAGGAGGCTCACCCACACTGCTCAATTTGTTGATGATCAAG 421
 QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAsp 661
 DB 422 ATGTGCTACTATCGTTTGGCAATATGATACGAGAGGTGGCAAGCCCGGCTACGAT 481
 QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGly 681
 DB 482 CGAGTTCTGCGCGGAGATCGCAACAAGGACTTTTGAACGTGGATGCTCTGGAGAGCG 541

301 AACATCCCCATATTGCTTCGTGTCTGAGCATCAGCCACACCTGGTCTCATACTAT 360

QY 362 PheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrCysPheSerAsn 381
 Db 361 TTTGACCTGGAGCTCTCTGCTTCATGTTTCCAGTTGGCTCTTATTACTGCTTAGCAAC 420
 QY 382 LeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyr 398
 Db 421 CTGCTGTATGCCCGGATTTTATCATCATCATGATGTTGGTGACCACTGATC 471

RESULT 6

US-09-248-796A-3089
 ; Sequence 3089, Application US/09248796A
 ; Patent No. 6747137

; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; PRIOR FILING DATE: 1999-02-12

; PRIOR FILING DATE: 1999-02-12
 ; PRIOR FILING DATE: 1998-02-13

; PRIOR FILING DATE: 1998-02-13
 ; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 3089

; LENGTH: 867
 ; TYPE: DNA

; ORGANISM: Candida albicans
 ; FEATURE:

; NAME/KEY: unsure
 ; LOCATION: (25)

; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unkno

US-09-248-796A-3089

Alignment Scores:

Pred. No.: 1,11e-84 Length: 867
 Score: 776.50 Matches: 147
 Percent Similarity: 76.13% Conservative: 38
 Best Local Similarity: 60.49% Mismatches: 53
 Query Match: 21.00% Indels: 5
 DB: 4 Gaps: 3

US-10-028-384-12 (1-705) x US-09-248-796A-3089 (1-867)

QY 460 MetIleLeuValMetAlaPheLeuIleThrTyrPheHisSerThrTrpValThr 479
 Db 52 GTTTTACTGCACATTTTACATTTTCTACTTTGTTTACATTTGTTTGGGTAAACA 111
 QY 480 SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArg 499
 Db 112 TCGAATGCTTATTCATCACCATCAGTTGTTTNGCATCCAGAAACCCAGATGGCTCACAA 171
 QY 500 IleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAsp 519
 Db 172 CATATCATTTGATATATAGAGGCTTACTGTTTGAAGTGAATGAATACACAGAGAT 231
 QY 520 AlaValValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaValAlaSerArgThr 539
 Db 232 GCCAAGATTATGGCTTGGGTGGGATTATGTTTATCAATCGGGGTATGGCTGATAGAACC 291
 QY 540 IleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMet 559
 Db 292 ACATCTGTTTGATAACAATACATGGAATAACACACATATTCACATGTTGGTGAAGCAATG 351
 QY 560 AlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeu 579
 Db 352 TCTTCCCTCCGAGATGTGCGTATGAATTTTGAGACACACAGATGTTGATTATGTGTA 411
 QY 580 ValIlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrpMet 599
 Db 412 GTTATATTTGGAGGTTATTTGGTTATCTGCTGATGATATTTAACAAATTTCTATGGATG 471

QY 600 ValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThr 619
 Db 472 GTAAGAATTTGCTGAAGGTATC--TGGCTGATGAATCAAAGAAAGAGACTACTTTACT 528
 QY 620 ProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMet 639
 Db 529 GACCGGAGAAATATAAGTGGATTAAGATGATCATCTGGCAATGAAGAATTTCTTTGATG 588
 QY 640 TyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGly 659
 Db 589 TATAAGTTATCGTATCATAGATTCACTGAATTTGTTGGAGGTAGA-----CATGGT 639
 QY 660 PheAspArgValArgAsnAlaGluIle--GlyAsnLysAspPheGluLeuAspValLeu 678
 Db 640 GTTGATAGAGTTAGAAACCAAAATCCAGCAATGAATGAATACCGAAATTTGAATTTGTT 699
 QY 679 GluGluGlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAsp 698
 Db 700 GAAGAAGCCTTCACATCAGAAATTTGGATTGTGAGANTTTACAAAGTTTAAAGATTGGAT 759
 QY 699 AsnArgGly 701
 Db 760 AATGTTGGT 768

RESULT 7

US-09-270-767-158

; Sequence 158, Application US/09270767
 ; Patent No. 6703491

; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 158

; LENGTH: 1660
 ; TYPE: DNA

; ORGANISM: Drosophila melanogaster
 ; US-09-270-767-158

Alignment Scores:

Pred. No.: 6.08e-84 Length: 1660
 Score: 774.50 Matches: 143
 Percent Similarity: 81.09% Conservative: 20
 Best Local Similarity: 71.14% Mismatches: 31
 Query Match: 20.95% Indels: 7
 DB: 4 Gaps: 2

US-10-028-384-12 (1-705) x US-09-270-767-158 (1-1660)

QY 498 SerArgIleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 517
 Db 788 TCCCGCAACATTTTAGACGATTTCCAGAGGCTTACTTGGCTTTCGCGAGACACTGCC 847
 QY 518 GluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsn 537
 Db 848 GATGATGCTCGCTTATGTTCTTGTGGATTACGATACACAGATAGCGGGAATGGCAAC 907
 QY 538 ArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGln 557
 Db 908 AGAACGACGCTAGTGGATATAATACGTGGAACAATAGTACATCTCTTGACGTGGACTAC 967
 QY 558 AlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyr 577
 Db 968 GCAATGCTTCAACCGGAGAGAGTCTCTACGAATTTATGACATCTCTTGACGTGGACTAC 1027
 QY 578 ValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeu 597
 Db 1028 GTTTTGTGATCTTTGGCGGTGTGATCGGCTATTCTGGCGATGATATCAACAAGTTCTTG 1087
 QY 598 TrpMetValArgIleGlyGlySerThrAspThrGlyLysHis-----IleLysGlu 614

Db 1088 TGGATGGTCCGAATTGCT-----GAGGAGAGCATCCCAAGGACATTAGGAA 1135
Qy 615 AsnAspTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeu 634
Db 1136 AGCGATTACTTACCGACCGGGTGAATTACAGGCTAGATCCGAAGGTGCTCCGGCCCTG 1195
Qy 635 LeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAla 654
Db 1196 CTCAACTGCCCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTAC 1255
Qy 655 LysArgProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGlu 674
Db 1256 AGAGTCCCATCTGGATATGTCGACACGTAAACCGCTATTGGGAATTAAGGACTTCGAT 1315
Qy 675 LeuAspValLeuGluGluGlyTyrThrGluHisTrpLeuValAlaGlyTyrLysVal 694
Db 1316 CTGACCTACCTGGAGAGGCTTACACACAGAACACTGGCTTGTTCGCATCTATAGGGTG 1375
Qy 695 Lys 695
Db 1376 AAG 1378

RESULT 8
US-09-270-767-15440
; Sequence 15440, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

Alignment Scores:
Pred. No.: 6,08e-84 Length: 1660
Score: 774.50 Matches: 143
Percent Similarity: 81.09% Conservative: 20
Best Local Similarity: 71.14% Mismatches: 31
Query Match: 20.95% Indels: 7
DB: 4 Gaps: 2

US-10-028-384-12 (1-705) x US-09-270-767-15440 (1-1660)
Qy 498 SerArgIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 517
Db 788 TCCCGCAACATTTAGACGATTTACAGAGAGGCTTACTAGCTTTCGAGAACACTGCC 847
Qy 518 GluAspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnIleAlaMetAlaAsn 537
Db 848 GATGATGCTCGCTTATGCTTGGTGGATTACGATACCATACGATACCGGATGCAAC 907
Qy 538 ArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGln 557
Db 908 AGAACGACGCTAGTGATATAATACGTGAACATAGTACATAGCGCTGCTGGCAAG 967
Qy 558 AlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyr 577
Db 968 GCAATGCTTCAACCGAGGAGAGTCTTACGAAATATGACATCTTTCAGCTGGACTAC 1027
Qy 578 ValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeu 597
Db 1028 GTTTTGGTATCTTGGCGGTGATCGGCTATCTTGGCGATGATATCAACAGTTCTCTG 1087
Qy 598 TrpMetValArgIleGlyGlySerThrAspThrGlyLysHis-----IleLysGlu 614

Db 1088 TGGATGGTCCGAATTGCT-----GAGGAGAGCATCCCAAGGACATTAGGAA 1135
Qy 615 AsnAspTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeu 634
Db 1136 AGCGATTACTTACCGACCGGGTGAATTACAGGCTAGATCCGAAGGTGCTCCGGCCCTG 1195
Qy 635 LeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAla 654
Db 1196 CTCAACTGCCCTTATGTACAAATTAAGCTACTACAGATTCCGGGAATTGAAGTTGGACTAC 1255
Qy 655 LysArgProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGlu 674
Db 1256 AGAGTCCCATCTGGATATGTCGACACGTAAACCGCTATTGGGAATTAAGGACTTCGAT 1315
Qy 675 LeuAspValLeuGluGluGlyTyrThrGluHisTrpLeuValAlaGlyTyrLysVal 694
Db 1316 CTGACCTACCTGGAGAGGCTTACACACAGAACACTGGCTTGTTCGCATCTATAGGGTG 1375
Qy 695 Lys 695
Db 1376 AAG 1378

RESULT 9
US-09-270-767-27262
; Sequence 27262, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 27262
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-27262

Alignment Scores:
Pred. No.: 1.13e-73 Length: 503
Score: 683.00 Matches: 127
Percent Similarity: 86.23% Conservative: 17
Best Local Similarity: 76.05% Mismatches: 23
Query Match: 18.47% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-270-767-27262 (1-503)
Qy 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
Db 2 GGTATATCTCTCGATCCGTCGGGGATCGTAGCAATGAAGGCATCGCCATTTCTGC 61
Qy 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
Db 62 ATGCTCTTCACTACTATTGTGGATCAAGGGGGTAAAGACGGGACGATCTTTGGTCG 121
Qy 196 AlalysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
Db 122 GCTATGTCGCATTTGGCTTACTTATATGGTCTCTCTCGGGGTGGCTATGTTCTCTG 181
Qy 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
Db 182 ATTAACCTAATCCCGTGCACGTGCTGGCGCTGATGATCACCGGACGTTCTTCGACAG 241
Qy 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
Db 242 ATCTACATGATACACACGCTTACTTACTGCTCGGACCATCTTCTGTCGATGACATCTCG 301
Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
Db 302 TTTGTGGATTCCAAACCATCCAGAGCTCCGAAACACATGCTGCGACTGGGAACCTTTGGC 361


```

Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
328 GTTCGCATCTATAGTGGAAG 308

RESULT 12
US-09-270-767-15440/c
; Sequence 15440, Application US/09270767
; Patent NO. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and
; FILE REFERENCE: File Reference: 7326-0
; CURRENT APPLICATION NUMBER: US/09/270,
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15440
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15440

```

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3090
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3090

Alignment Scores:
Pred. No.: 1,948-59 Length: 594
Score: 567.50 Matches: 103
Percent Similarity: 74.86% Conservativeness: 31
Best Local Similarity: 57.54% Mismatches: 44
Query Match: 15.35% Indels: 1
DB: 4 Gaps: 1

US-10-028-384-12 (1-705) x US-09-248-796A-3090 (1-594)
QY 7 LeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeuLeuLeuSerMetAla 26
Db 58 TTGGGGATAGATGTTGAACGCTAGTATTAGAGTTTATATAAAGTTATATATATATCGATA 117
QY 27 AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis 46
Db 118 GCAGGTGCAGCTATTCCTTCGTTATTTCCGCGATTCGATTTGAAGATTATTCAT 177
QY 47 GluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluGlyPheTyr 66
Db 178 GAATTCGATCCCTGTGTTCAATTTCCGACCAACCAATATATTTAGTCACATTCCTTTTAT 237
QY 67 LysPheHisAsnTrpPheAspArgAlaTrpTyrProLeuGlyArgIleIleGlyGly 86
Db 238 GAATTTTCAATTTGTTGATGATAGAACTTGGTACCCATTTGGGAGAGTCACCTGGTGT 297
QY 87 ThrIleTyrProGlyLeuMetIleThrSerAlaIleTyrHisValLeuHis---Phe 105
Db 298 ACTTATATATCCCGTTTAATGGTGACTTCAGGTGCCATTTGGCATATTTTACGTGATGG 357
QY 106 PheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerSer 125
Db 358 TTTCCTTACCCGTTGATATATAAATTTGTGTTTATTAGCACCCAGTTTCTCGGA 417
QY 126 PheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyAlaGlyLeu 145
Db 418 TTAACCTGCAATTTGACTATATTTTGTACTAAGAAATGAACGATTTCTAGTGCACGATTA 477
QY 146 LeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArgSerValAlaGlySer 165
Db 478 TTGGCAGCTATATTTATCGGGATTGCCCCAGGTATATTTCAAGATCAGTGGCTGTCT 537
QY 166 TyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrMetTrpIle 184
Db 538 TATGATAATGAAGCAATTTGCCATTACTTTATTAATGAACACATTTATTTCTGGATT 594

RESULT 14
US-09-702-705-1655
; Sequence 1655, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedwick, Tom
; APPLICANT: Carter, Darrick

```

; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1655
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-1655

Alignment Scores:
Pred. No.: 1,42e-57 Length: 487
Score: 551.00 Matches: 104
Percent Similarity: 91.53% Conservative: 4
Best Local Similarity: 88.14% Mismatches: 0
Query Match: 14.90% Indels: 10
DB: 4 Gaps: 1

US-10-028-384-12 (1-705) x US-09-702-705-1655 (1-487)

```
QY 477 TrpValThrSerGlu-----AlaTyrSerSerPro 486
Db 29 TGGATATCTCGAGAAATTCGCCCTTTTCGAGCGCGCGCCGGCAGGTCCTACTCTCTCCG 88
QY 487 SerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspPheArg 506
Db 89 TCCATTGTACTATCTGCCCGGTGGGGATGGCAGTAGGATCATATTTGATGACTTCCGA 148
QY 507 GluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrp 526
Db 149 GAAGCATATATTGGCTCCGTCAATAACTCCAGAGGATGCGAAGGTCATCTCTGTGG 208
QY 527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThr 546
Db 209 GATTATGGCTATCAGATTACAGTATGCGAAACCGAACAAATTTAGTGACCAATAACACA 268
QY 547 TrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAla 566
Db 269 TGGAAATAATACCCATATTTCTCGAGTAGGGCAGGCAATGGCGTCCACAGAGGAAAAGCC 328
QY 567 TyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGly 584
Db 329 TATGAGATCATGAGGAGCTCGATGTCAGTATGCTGTGTCATTTTGGAGGA 382
```

RESULT 15

US-09-736-457-1655
; Sequence 1655, Application US/09736457
; Patent No. 6509448

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1655

; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-736-457-1655

Alignment Scores:
Pred. No.: 1,42e-57 Length: 487
Score: 551.00 Matches: 104
Percent Similarity: 91.53% Conservative: 4
Best Local Similarity: 88.14% Mismatches: 0
Query Match: 14.90% Indels: 10
DB: 4 Gaps: 1

US-10-028-384-12 (1-705) x US-09-736-457-1655 (1-487)

```
QY 477 TrpValThrSerGlu-----AlaTyrSerSerPro 486
Db 29 TGGATATCTCGAGAAATTCGCCCTTTTCGAGCGCGCGCCGGCAGGTCCTACTCTCTCCG 88
QY 487 SerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspPheArg 506
Db 89 TCCATTGTACTATCTGCCCGGTGGGGATGGCAGTAGGATCATATTTGATGACTTCCGA 148
QY 507 GluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrp 526
Db 149 GAAGCATATATTGGCTCCGTCAATAACTCCAGAGGATGCGAAGGTCATCTCTGTGG 208
QY 527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThr 546
Db 209 GATTATGGCTATCAGATTACAGTATGCGAAACCGAACAAATTTAGTGACCAATAACACA 268
QY 547 TrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAla 566
Db 269 TGGAAATAATACCCATATTTCTCGAGTAGGGCAGGCAATGGCGTCCACAGAGGAAAAGCC 328
QY 567 TyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGly 584
Db 329 TATGAGATCATGAGGAGCTCGATGTCAGTATGCTGTGTCATTTTGGAGGA 382
```

Search completed: December 15, 2004, 12:00:08

Job time : 165.129 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 15, 2004, 06:02:06 : Search time 887.605 Seconds
(without alignments)
4379.236 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKGFLRLSEKQDTLLKL.....HMLVRIYKVDLNRGLSRT 705

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4085731 seqs, 2756760397 residues

Total number of hits satisfying chosen parameters: 8171462

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-Q=/cgn2_1/USPTO_spoos/US10028384/runat_14122004_131523_8140/app_query.fasta_1.3740
-DB=Published Applications NA -GFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blcostum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10028384@cgn1.1.2577 -runat_14122004_131523_8140
-NCPU=6 -ICPU=3 -NO WAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRAIDS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 3697 | 100.0 | 2472 | 15 | US-10-171-581-112 |
| 2 | 3697 | 100.0 | 2472 | 15 | US-10-028-384-11 |
| 3 | 3697 | 100.0 | 2472 | 15 | US-10-172-118-742 |
| 4 | 3697 | 100.0 | 2472 | 16 | US-10-342-887-742 |
| 5 | 3657 | 98.9 | 3046 | 18 | US-10-417-375-95 |
| 6 | 3633 | 98.3 | 3093 | 18 | US-10-417-375-92 |
| 7 | 3633 | 98.3 | 3094 | 15 | US-10-028-384-9 |
| 8 | 3597 | 97.3 | 3404 | 18 | US-10-417-375-99 |
| 9 | 3597 | 97.3 | 3404 | 18 | US-10-417-375-97 |
| 10 | 3258 | 61.1 | 2839 | 18 | US-10-425-115-150745 |
| 11 | 3258 | 61.0 | 2839 | 18 | US-10-425-115-150745 |
| 12 | 3247 | 60.8 | 2481 | 16 | US-10-425-114-14408 |
| 13 | 3237 | 60.5 | 2710 | 15 | US-10-028-384-1 |
| 14 | 3237 | 60.5 | 2710 | 15 | US-10-028-384-3 |
| 15 | 2177.5 | 58.9 | 2417 | 15 | US-10-437-963-99904 |
| 16 | 2070.5 | 56.0 | 2256 | 15 | US-10-028-384-7 |
| 17 | 2010 | 54.4 | 2232 | 15 | US-10-032-585-6323 |
| 18 | 1940.5 | 52.5 | 3141 | 18 | US-10-128-714-7139 |
| 19 | 1930 | 52.2 | 2466 | 16 | US-10-425-115-130787 |
| 20 | 1929 | 52.2 | 2157 | 9 | US-10-320-797-2305 |
| 21 | 1929 | 52.2 | 2157 | 9 | US-09-801-368-387 |
| 22 | 1920 | 51.9 | 2733 | 15 | US-10-793-639-318 |
| 23 | 1919 | 51.9 | 2733 | 15 | US-10-028-384-5 |
| 24 | 1851 | 50.1 | 2603 | 15 | US-10-424-595-111541 |
| 25 | 1851 | 50.1 | 3969 | 15 | US-10-128-714-6139 |
| 26 | 1851 | 50.1 | 3969 | 15 | US-10-128-714-139 |
| 27 | 1722 | 46.6 | 1848 | 15 | US-10-128-714-5139 |
| 28 | 1682 | 45.5 | 2882 | 16 | US-10-128-714-2139 |
| 29 | 1682 | 45.5 | 4738 | 16 | US-10-320-797-1305 |
| 30 | 1680 | 45.4 | 1969 | 15 | US-10-320-797-305 |
| 31 | 1651.5 | 44.7 | 2244 | 17 | US-10-128-714-1139 |
| 32 | 1336.5 | 36.2 | 1728 | 16 | US-10-437-963-48342 |
| 33 | 1241.5 | 33.6 | 1209 | 15 | US-10-424-599-122476 |
| 34 | 1241.5 | 33.6 | 1209 | 15 | US-10-106-698-330 |
| 35 | 1241.5 | 33.6 | 1209 | 16 | US-10-264-237-412 |
| 36 | 1241.5 | 33.6 | 1543 | 10 | US-09-974-879-133 |
| 37 | 1241.5 | 33.6 | 1543 | 10 | US-09-305-736-133 |
| 38 | 1241.5 | 33.6 | 1543 | 11 | US-09-818-683-133 |
| 39 | 1241.5 | 33.6 | 1543 | 11 | US-09-818-683-133 |
| 40 | 1097.5 | 29.7 | 1094 | 17 | US-10-621-401-133 |
| 41 | 931 | 25.2 | 1828 | 10 | US-09-945-527-62 |
| 42 | 875.5 | 23.7 | 2660 | 16 | US-10-264-049-630 |
| 43 | 820.5 | 22.2 | 1114 | 16 | US-10-296-115-629 |
| 44 | 805 | 21.8 | 487 | 10 | US-09-871-161-213 |
| 45 | 802 | 21.7 | 452 | 9 | US-09-878-178-1456 |

ALIGNMENTS

RESULT 1
US-10-171-581-112
; Sequence 112, Application US/10171581
; Publication No. US2003010426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 112
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: L38961

! DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-112

Alignment Scores:

Pred. No.: 0 Length: 2472
Score: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-171-581-112 (1-2472)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
Db 107 ATGACTAAGTTGGATTTTGGGATTTCTATATGAGAGCAGACACACTTTTGAAGCTT 166
Qy 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGTCATATGGCTGCTGATTATCTCTCCACCTGCTGTTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTTGAAAGTGATPCCAGATGATGATCGTACTTTTAAATATCGGACTACCAAGGTTCCTG 286
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrPyrProLeu 80
Db 287 GCTCAGGAGGGGTTTATAAATCCATACTGGTGTGATGACCGAGCGCTGGTACCCTTTG 346
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGAGCATCATGTGGAGAACATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTCCACATCACCATCGACATCGGAATGTCTGTGTGTTCTCGGCC 466
Qy 121 ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTTCTGTGCTGGCATGATGCTGTAGTTCCTGGATATATCTCCCGA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATAATGAGGATTTGCCATCTTTTGCATGCTACTCACCTAC 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGGCAGTAAAGACTGGTTCCATCTGTGGGCAGCTAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 707 GCTTATTTCTACATGGTCTCTCATGGGAGGTTATGTTCTCTGATCACTTAATCTCT 766
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACATCTCTGCTGCTGATGCTCACAGGCCGCTTTCTCTCACCGGATCTATGTGGCTTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTAATGTTTACTGCTGGTATCTATATCTTTCTAGSCAGATCTCTTTGTGGGTTTCCAG 886
Qy 261 ProValLeuSerGlnHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CCTGCTCTTTCTATCAGACATATGGCAGGCTTTGGGCTCTTTGGTCTCTGCGCAGATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db 947 GCCTTTGTGGATACCTGGCAGCAGTGTGATCCACACAAATTTGAAGTCTTTTCCGG 1006

Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGCTAGGCTTTTGTCTCTCACCGTGGGAGCTCTCTCTCATGCTGACA 1066
Qy 321 GlyValIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1067 GGAATAAATATCTCCCTGGACGGGCGTTCCTACTCTGCTGGATCCCTCTTAAGTAAAG 1126
Qy 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1127 AACAAACATCCCATCATTGCTTCTGTGTCTGAGCATCAGCCCAACACTGGTCTCTCATAC 1186
Qy 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1187 TATTTTGGCTGGACCTCTCTCTCTTATGTTTCCAGTTGGCTCTTACTTACTGCTTAGC 1246
Qy 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
Db 1247 AACCTGTCTGATGCCCGGATTTTATCATCATGATGATGCTGTCAGCAGCATGTACTTTTCA 1306
Qy 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1307 GCTGTAAATGGTGGCTCTAATGCTAGTGTGGACCTGTTATGAGCATCTCTCTGSCATT 1366
Qy 421 GlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProAspLys 440
Db 1367 GGAGTCTCCAGGTGCTGTCCACATACATGAAGAATCTGGACATAAGTCGCCGACACAAG 1426
Qy 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
Db 1427 AAGACAAAGAACAAACAGGATTCACCTACCTATTAAGATTGAGTGGCAAGTGGGATG 1486
Qy 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
Db 1487 ATACTGGTCATGGCTTCTTCTCATCACCTACACCTTTCATTCACCTGGGTGACCAT 1546
Qy 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1547 GAGGCTACTCTCTCCGCTCTGTTACTATCTGCCGCTGTTGGGATGGCAGTGGATC 1606
Qy 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAlaAla 520
Db 1607 ATATTGTGATGATTCGAGAAAGCATATATTGGTTCGTCATATACTCTCAGAGGATGG 1666
Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGTTCATGTCCTGGTGGATTTGCTATCAGATTACAGTATGTCNAACCGAACAT 1726
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTAGTGGACAATAACATGGAATAATACCCATATTTCTCAGTAGGCGAGCAATGGCG 1786
Qy 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGGAAAAGCCATGAGATCATGAGGAGCTCGATGTGACGATATGCTGCTGTC 1846
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCCTGCGGTATTTCTCTGATGATATCAACAGTTTCTTTGGATGGTC 1906
Qy 601 ArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTTGGGAGGACACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCA 1966
Qy 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGTTCGCTGGTGGACCGCTGAAAGGTTCCTCAGTGTGCTCAACTGCCCTCATGTAC 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2027 AAGATGTGTACTCTGCTTTGGAGAGGTTTACACAGAACCAAGCGTCTCTCCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680


```

Db 1607 ATATTCATGACCTCCGAGAGCATATATTGGCTTCGTCAATAATCCAGGATCGG 1666
Qy 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1667 AAGGTCATGCTCCGTGGGATTATGGCTATCAAGATTACAGCTATGGCAACCGACAAT 1726
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1727 TTATGGACAAATACACATGATATATACCATATTTCTCGAGTAGGACGAGCATGGCG 1786
Qy 561 SerThrGluGluLysAlaTyrGlyIleMetArgGluLeuAspValSerTyrValLeuVal 580
Db 1787 TCCACAGAGAAAGGCTATGAGATCATGAGGAGCTCGATGTGCTATGCTGCTC 1846
Qy 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1847 ATTTTGGAGGCTCCTGCTGATTTCTCTGATGATATCAACAAGTTCTTTGGATGGTC 1906
Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
Db 1907 CGGATTGGAGGAGCACAGATACAGGCAAAACATATCAAGGAGATGACTATTATCTCCA 1966
Qy 621 ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1967 ACTGGGAGTTCGGTGGGACCTGAGGTTCTCCAGTCTGCTCAACTGCTCATGTAC 2026
Qy 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2027 AAGATGTGTTACTATCGCTTTGGACAGGTTTACACAGAAGCCAGCGCTCTCCAGGCTTT 2086
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2087 GACCGTCTCGGAATGCTGAGATGGGAATTAAGACTTTGAGCTTGATGTCTCTGGAGAA 2146
Qy 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2147 GGCTATACCAAGAACATTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2206
Qy 701 GlyLeuSerArgThr 705
Db 2207 GGCTTTGTCAAGGACA 2221

```

RESULT 3

```

US-10-172-118-742
; Sequence 742, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mac, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_002219
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-742

```

Alignment Scores:

```

Pred. No.: 0 Length: 2472
Score: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservations: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-172-118-742 (1-2472)

Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTATGATTGGGATTTTTCGATTTCCTATGAGAGCAGGACACACTTTTGAAGCTT 166
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATTCTGCTCAATGGCTGCTGATTATCTCTCCACTGCTGTTGTTGCTGCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTGGAAGGTTTATCCATGAGTTTGCATCCGTTTAAATATCGAGTACAGGTTCTCTG 286
Qy 61 AlaGluGlyGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaIleTyrProLeu 80
Db 287 GCTGAGGAGGGGTTTATAAATTCATAAATCGTTTGCATGCCGAGGCTGGTACCCCTTG 346
Qy 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 347 GGACGATCATTTGGAGGAACAATTTTACCCAGGTTTAAATGATCACCTCTGTCGAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCATCGACATCGAATCGGAATGCTGTGTGCTTGGCC 466
Qy 121 ProLeuPheSerSerPheThrSerIleValTyrThrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CTTCTCTTCTCTCTCTCTCACCTCCATGTCACGTCACCTCTCTACCAAGAGCTCAAGAT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTTCTTGTCTGCTGCCATGATGCTGTAGTTCCTGGATATATCTCCGA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGGGCTGGCTCTCTATGATTAATGAAGGATTCGCAATCTTTTGCATGCTACTCACTAC 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGGACGATGAAGACTGGTTCCATCTGTTGGGAGGCTAAGTGTGCCCT 706
Qy 201 AlaTyrPheTyrMetValSerSerTrpGlyClyTyrValPheLeuIleAsnLeuIlePro 220
Db 707 GCTTATTTCTACATGCTCTGTCATGGGAGGTTATGTGTTCTTGATCACTAATCTT 766
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 767 CTCACGTCCTGCTGCTGATGCTCACAGGCGGTTTCTCTACCGGATCTATGTGGCTAC 826
Qy 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 827 TGTACTGTTTACTGCTGGGACTATATCTTCTAGGAGATCTCTTGTGGGGTTTCCAG 886
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 887 CCTGTCTCTTATCATCAGACACATGGCAGGGTTGGGGTCTTTGGTCTCTGCCAGATCCAT 946
Qy 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
Db 947 GCCTTTGTGGATTACTGCGGACGACAGTTGAATCCACACAAATTTGAAGTCTTTCCGG 1006
Qy 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1007 AGCGTCATCTCTCTGTTAGGCTTTTGTCTCTCACCCTGGGAGCTCTCTCTCATGCTGACA 1066

```

QY 321 GlyIysIleSerProThrThrGlyArgPheTy-SerLeuLeuAspProSerTyrAlaLys 340
DB 1067 GAAAAATATCTCCCTGACGGGCGTTTCTACTACTGCTGGATCCCTCTTATGTAAG 1126
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrThrPheSerTyr 360
DB 1127 AACAAATCCCATCAATTCCTGCTGAGCATCAGCCCAACCTGGTCTCATAC 1186
QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
DB 1187 TATTTTGACCGGAGCTCCCTGCTTCTCATGTTTCCAGTTGGCTCTTATCTAGCTTTAG 1246
QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
DB 1247 AACCTGTCTGATGCCCGGATTTTATCATCATGTATGTTGACCATGACTTTTCA 1306
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
DB 1307 GCTGTAAATGGTGGCTTAAGTCTAGTGTGGACCTTGTATGACATCTCTCTGGCAT 1366
QY 421 GlyValSerGlnValLeuSerThrTyrMetIleAsnLeuAspIleSerArgProAspLys 440
DB 1367 GGAGTCTCCAGGTGCTGCACATACATGAAGAATCTGCACATAAGTCCGCCAGACAAG 1426
QY 441 LysSerIleGlnGlnAspSerThrTyrProIleIleGlnValAlaSerGlyMet 460
DB 1427 AAGAGCAAGAACCAAGGATTCACCTACCTTATTAAGATTGAAGTGGCAAGTGGGATG 1486
QY 461 IleLeuValMetAlaPheLeuIleThrTyrPheHisSerThrThrValThrSer 480
DB 1487 ATACTGGTCAATGGCTTCTTCTCATCACCTACACCTTCTCATCAACCTGGGTGACCAT 1546
QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
DB 1547 GAGGCCACTCTTCTCGCTCATTTGTAATCTATCTGCGCGTGGTGGGATGCGATGAGATC 1606
QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
DB 1607 ATATTTGATGATCTCCGAGAGCATATATTGGCTTCTCATATAATCTCCAGGATGCG 1666
QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaSerThrIle 540
DB 1667 AAGGTCATGCTCTGGTGGATATGGCTATCAGATTACAGCTATGCGCAACCGCAAT 1726
QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
DB 1727 TTAGTGGACAATAACATGAATAATACCATATTTCTGAGTAGGCGAGCAATGGG 1786
QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
DB 1787 TCCACAGAGGAAAGCCCTATGAGATCATGAGGAGCTCGATGCTAGCTATGCTGGTC 1846
QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
DB 1847 ATTTTGGAGCCCTCACTGGGTATTCCTCTGATGATATCAACAAGTTCTTTGGATGTC 1906
QY 601 ArgIleGlyGlySerThrAspThrGlyHisIleIleGluAsnAspTyrThrPro 620
DB 1907 CGGATTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCCA 1966
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
DB 1967 ACTGGGAGTTCCGTGGACCGGTGAAGTTCTCCAGTGTCTCAACTGCCTCATGTAC 2026
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
DB 2027 AAGATGTCTTACTATCGCTTTGGACAGGTTTACACAGAACCAAGCGTCTCCAGGCTTT 2086
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
DB 2087 GACCGTGTCCGAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGTATGCTCTGGAGAA 2146
QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700

DB 2147 GGCTATACCAGAAATTCGCTGGTCAGATATACAGTAAGACCTGGATATCA 2206
QY 701 GlyLeuSerArgThr 705
DB 2207 GGCTTGTCAAGGACA 2221
RESULT 4
US-10-342-887-742
; Sequence 742, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 742
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-742
Alignment Scores:
Pred. No.: 0 Length: 2472
Scores: 3697.00 Matches: 705
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-10-028-384-12 (1-705) x US-10-342-887-742 (1-2472)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
DB 107 ATGACTAAGTTTGATTTTTCGATTTCTATGAGAACGACGACACATTTTGAAGCTT 166
QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 167 CTCAATCTGTCAATGGCTGTGTATTATCTTCTCCACTCGTCTGTGCTGCTGAGA 226
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 227 TTTGAAAGTGTATCCATGAGTTTATCGTACTTTTATATTCGGACTACCAAGTTCCG 286
QY 61 AlaGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
DB 287 GCTGAGGAGGGTTTATAAATCCATACTGGTTTGTATGACCGAGCTGTACCTTTG 346
QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 347 GGACGAATCATTTGAGGAACAATTTACCAGGTTTAAATGATCACCTCTCTGCAATCTAC 406
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 407 CATGCTACTCCATTTTCCACATCAATCGACATTCGGAATGCTGTGTGTTCTGCTGCC 466
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140

US-10-417-375-92
 ; Sequence 92, Application US/10417375
 ; Publication No. US20040219528A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 ; FILE REFERENCE: 529452001600
 ; CURRENT APPLICATION NUMBER: US/10/417,375
 ; CURRENT FILING DATE: 2003-04-15
 ; NUMBER OF SEQ ID NOS: 176
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 92
 ; LENGTH: 3093
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; US-10-417-375-92

Alignment Scores:
 Pred. No.: 0 Length: 3093
 Score: 3633.00 Matches: 694
 Percent Similarity: 98.87% Conservatives: 3
 Best Local Similarity: 98.44% Mismatches: 8
 Query Match: 98.27% Indels: 0
 DB: 18 Gaps: 0

US-10-028-384-12 (1-705) x US-10-417-375-92 (1-3093)

QY 1 MetThrIysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 112 ATGACTAAGCTTGATTTTGGCATTTGCTTATGAGAGCAGACACACTTCTAAAGCTT 171
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 172 CTCATCCTGCGATGGCTGCTGCTGTTATCTTTTCTACTCGTCTTTTCTGCTGCTGAGA 231
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 DB 232 TTTGAAAGTGTCATCAATGATTTGATTCGCTACTTTTAAATATCGGACTACCGGTTCTG 291
 QY 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeu 80
 DB 292 GCTGAGGAGGGTTTATAAATCCATAACTGTTTATGATGACGGGCTGGTACCCCTTG 351
 QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 352 GCGCGAATCAITGGAGAACAAATTTACCAGGTTTAAATGATCACTTCTGCTGCAATCTAC 411
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 412 CATGTACTCCATTTCTCCATATCACTAATGACATTCGGATGTCGTGTTTCTGCGCC 471
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 DB 472 CCACCTTTCCTCTTCCACCATCGTTAGTACACCTTACCACAGAGCTCAAGGAT 531
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
 DB 532 GCAGGAGCTGGGCTTCTTGCTGCTGCATGATTTGCTGATTTCTGCGGTATATTTCGA 591
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 DB 592 TCTGTAGTGGCTTCTATGATAAAGGAATTTGCTATCTTTTGTGCTGCTGCTTACTTAC 651
 QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
 DB 652 TACATGTGGATCAAGGACGAGTGGTTCATCTATTGGGCTGCCAAGCTGCGCCTC 711
 QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
 DB 712 GCATTATTTCTACATGTCCTCTTATGGGAGGCTATGCTGCTGCTGCTGCTGCTGCT 771
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240

DB 772 CTACATGTCCTGGTCTAATGCTGACAGCGCGTTTTTCTCACCGGATCTACGTAGCTAC 831
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 832 TGTACTGTTTACTGCTGGGACCATTTCTTCTATGAGATTTCCCTTTGTTGGTTCCAG 891
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 892 CCGCTCTTTTCTATCAGAACACATGCGAGCCTTTGGAGTGTGTTGCTCTCTCTCAGATCCAT 951
 QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGlnValLeuPheArg 300
 DB 952 GCTTTCTGTAGATTACTGCGCAGCAAGTTGAATCCACAGCAATTCGAGATCTCTTTCCGG 1011
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 DB 1012 AGTGTATCTCCCTGGTTGGCTTTGCTCTCTCTACTGCGGAGCTCTCTCTCATGTAACA 1071
 QY 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 DB 1072 GGAATAATTTCTCCCTGACAGGCGTTTCTACTCTCTGCTGATCCCTCTTTATGCTAAG 1131
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
 DB 1132 AATAACATTTCCATTTATGTCATCTGTTTCTGAGCAGCAGCCACCACTGGTCTCTCTAC 1191
 QY 361 TyrPheAspLeuLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1192 TATTTTGTACTACAGCTCTGCTCTCTCATGTTCCAGTTGCTGCTCTATTTACTGCTTACG 1251
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1252 AACCTGTGTGATGCTCGGATTTTATCATCATGATGATGATGATGATGATGATGATGAT 1311
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1312 GCTGTATGTTGGTGGCTCTAATGCTGGTATGTCACCTGTTATGTCATCTCTTTCTGGCAT 1371
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 DB 1372 GGTGTCTCCAGTCTGTCACATATATGAAAAATCTGGACATAGTCCGCCAGACAAG 1431
 QY 441 LysSerLysLysGlnAspSerThrTyrProIleLysIleGluValAlaLysArgGlyMet 460
 DB 1432 AAGACAGAGACACAGATTTCTATCCCTATTAGATGAGTGGCGAGTGGGATG 1491
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1492 ATACTGGTTCATGGCTTTTCTCTCATCACCTACACGTTTCTATTCGAGTGGGACAGT 1551
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1552 GAAGGCTATTCTTCTCCCTCCATTTGACTGCTGCTGCTGGGATGCGAGTGGATC 1611
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1612 ATTTTGTATGATCTCCGAGAGACGCTATTATTGGTCTCGTCACATACTCCAGAGGATGCA 1671
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1672 AAGTCATGTCATGTTGGGATTTATGGCTACCAATTTACTGCAATGGCAATTCGACAT 1731
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1732 TTAGTGGACAATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAGCAATGGCA 1791
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1792 TCCACAGAGAGAAAAGCCTATGAAATCATGAGGAGCTTGTATGTCAGCTATGCTGCTG 1851
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600

Db 1852 ATTTTGGAGGCTTACTGGGTAATCTTCGGATGATATCAACAAGTTCTTTTGGATGGTC 1911
QY 601 ArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrPro 620
Db 1912 CGCAITGGAGGAAGACAGACAGACAGACACATTAAGGAGATGACTACTATCTCT 1971
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
Db 1972 ACTGGGAATTCGGTGTGATCGTGAGGGTCTCCGGTGCTGCTCAACTGCCCTTATGTAC 2031
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
Db 2032 AAAATGCTGTACTACCGCTTTGGGCGAGCTTACACAGAGCAAGCGTCACACAGGCTTT 2091
QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
Db 2092 GACCGTGTTCGAATCTGAGATGGTGAATTAAGACCTTGAGCTTGATGCTCGGAGGA 2151
QY 681 GlyTyrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2152 GCGTATACACAGACACTGGCTAGTCAGGATATACAAAGGTAAAGGACCTGGATATCGA 2211
QY 701 GlyLeuSerArgThr 705
Db 2212 GCGTGTCAAGGACA 2226

RESULT 7

US-10-028-384-9
; Sequence 9, Application US/10028384
; Publication No. US20030148285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAULT, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 9
; LENGTH: 3094
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 008408
; DATABASE ENTRY DATE: 2000-11-01
; RELEVANT RESIDUES: (1)..(3094)
US-10-028-384-9

Alignment Scores:

Pred. No.: 0 Length: 3094
Score: 3633.00 Matches: 694
Percent Similarity: 98.87% Conservative: 3
Best Local Similarity: 98.44% Mismatches: 8
Query Match: 98.27% Indels: 0
DB: 15 Gaps: 0

US-10-028-384-12 (1-705) x US-10-028-384-9 (1-3094)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
Db 112 ATGACTAAGCTTGGATTTTCGGATTGCTCTATGAGAAGCAGGACACACTTCTAAAGCTT 171
QY 21 LeuIleLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 172 CTCATCTCTGTCATCGCTGCTGTGTATCTTTTCTACTCGTCTTTTGTGCTGCTGAGA 231
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 232 TTTGAAAGTGTTCATCCATGAGTTGATCCGCTACTTTTAATATCGGACTACCGGTTCTG 291
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu 80

Db 292 GCTGAGGAGGGTTTTATAAAATCCATACTGTTTTGATGACGGGCTTGGTACCCCTTTG 351
QY 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
Db 352 GCGCGAATCATTTGAGGAGAACATTTACCCAGGTTTAAATGATCACTTCTGCTGCAATCTAC 411
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 412 CATCTACTCCATTTCTTCCATATCACTATTGACATTCGGAATGCTGTGTTTTCTCTGGCC 471
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 472 CCACCTTTCTCTCTTTCACCAACCATCTGATACCATCTTACCAAAAGAGCTCAAGAT 531
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 532 GCAGAGAGCTGGCTTCTTGTCTGCTGCATGATGCTGTAGTTCTCTGGGTATATTTCTCGA 591
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 592 TCTGTAGCTGGCTCTCTATGATATGAAGAAATGCTATCTTTTGTGATGCTGCTTACTTAC 651
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 652 TACATGTGATCAAGGAGCTGAAGACTGGTTCCATCTATTGGGCTGCCAAGTGTGCCCTC 711
QY 201 AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIlePro 220
Db 712 GCTTATTTCTACATGGTCTCTTTCATGGGAGGCTATGTTCTCTGATCAACTGATTCTCT 771
QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db 772 CTACATGCTCTGGTGCTTAATGCTGACAGCGCGTTTTTCTCACCGGATCTACGTAGCCTAC 831
QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 832 TGTAAGTGTACTGCTGGCGGACCATCTTCTATGACAGATTTCTTCTGTTGGTTCAG 891
QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 892 CCCGTCTCTTCATCAGAACACATGGGAGGCTTTGGAGTGTGTTGCTCTCTGTCAGATCCAT 951
QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db 952 GCTTTTGTAGATTACCTGCGCAGCAAGTTGAATCCACACCAATTCGAAGTCTTTTCCGG 1011
QY 301 SerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db 1012 AGTCTTATCTCCCTGGTGGCTTTGCTCTCTCCTCCTCCTGAGGAGCTCTCTCATGTAACA 1071
QY 321 GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db 1072 GGAAATTTCTCCCTGGACAGGCGGTTTCTACTCTCTGCTGATCCCTCTTATGCTAAG 1131
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db 1132 AATAACATTTCCCATTTATGCACTGTTTCTGAGCACCAGCCACACCTGGTCTTCTCTAC 1191
QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
Db 1192 TATTTTGTATGATACAGCTCTCTTGTCTTCAATGTTTCCAGTTGGCTCTTATCTGCTTAGC 1251
QY 381 AsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
Db 1252 AACCTGTCTGATGCTCGGATTTTATCATCATGATGGTGGTACCCAGCATGACTTTTCA 1311
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
Db 1312 GCTGTATGCTGCTCTAATGCTGGTATTTGGACCTGTTATGTCGATTTCTTCTGCGCAT 1371
QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440


```
Db 1372 GGTGTTTCCAGGTCGTCCACATATATGAAATCTGGACATTAAGTCGCCACAG 1431
Qy 441 LysSerLysGlnGlnAspSerThrThrProIleLysIleGluValAlaSerGlyMet 460
Db 1432 AAGAGCAAGAACACAGAGATTCTACTTACCTATTAAAGATGAGGTGGCAGTGGATG 1491
Qy 461 IleLeuValMetAlaPhePheLeuIleThrThrPheHisSerThrTrpValThrSer 480
Db 1492 ATACTGTCATGGCTTTTTCATCACCTACACGTTTCTCCACTTGGTGGTACCAGT 1551
Qy 481 GluAlaThrSerProSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
Db 1552 GAAGCTTATCTTCTCCCTCCATTTACTGTCTGCTGCTGGTGGGATGGCAGTAGGATC 1611
Qy 501 IlePheAspAspPheArgGluAlaThrTrpLeuArgHisAsnThrProGluAspAla 520
Db 1612 ATTTTGTATGACTCCGAGAAAGGTATTATTGGCTCCGTCACAACTCCAGAGATGCA 1671
Qy 521 LysValMetSerTrpTrpAspThrGlyThrGlnIleThrAlaMetAlaAsnArgThrIle 540
Db 1672 AAAGTCATGTCATGGTGGGATTTATGGCTACCAAAATTAATGCAATGCAAAATCGGACAA 1731
Qy 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
Db 1732 TTAGTGACAAATTAACATGGAATTAATACCATATTCTCGAGTAGGAGCAGGCAATGGCA 1791
Qy 561 SerThrGluGluLysAlaThrGluIleMetArgGluLeuAspValSerThrValLeuVal 580
Db 1792 TCCACAGAAAGAAAGCCTATGAATCATGAGGAGCTTGATGTCAGCTATGTGCTTGTGTC 1851
Qy 581 IlePheGlyGlyLeuThrGlyThrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
Db 1852 ATTTTGGAGGCTTACTGGGTATTCTTGGATGATATCAACAGATTTCTTTGGATGTC 1911
Qy 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspThrThrPro 620
Db 1912 CGGATTGGAGAGACACAGACAGAGAGGAGACACATTAAGCAGAAATGACTACTATCTCT 1971
Qy 621 ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetThr 640
Db 1972 ACTGGGAAATCCGTGTGATGCTGAGGTTCTCGGTGCTCTCACTGCTTATGTATC 2031
Qy 641 LysMetCysThrThrArgPheGlyGlnValThrGluAlaLysArgProProGlyPhe 660
Db 2032 AAAATGTGTTACTACCGCTTTGGGACAGGCTACACAGAAAGCCAAAGCGTCCACAGGCTTT 2091
Qy 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
Db 2092 GACCGTGTTCGAATGCTGAGATTTGGTAATTAAGACTTTGAGCTTGATGCTCTGGAGGAA 2151
Qy 681 GlyThrThrGluHisTrpLeuValArgIleThrLysValLysAspLeuAspAsnArg 700
Db 2152 GCGTATACCAAGAACACTGGCTAGTACAGATATACAGGTAAAGGACCTGGATAATCGA 2211
Qy 701 GlyLeuSerArgThr 705
Db 2212 GCCTTGTCAAGGACA 2226
```

RESULT 8

```
US-10-417-375-99
; Sequence 99, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 99
; LENGTH: 5404
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-417-375-99
Alignment Scores:
Pred. No.: 0 Length: 5404
Score: 3597.00 Matches: 686
Percent Similarity: 99.13% Conservatives: 1
Best Local Similarity: 98.99% Mismatches: 6
Query Match: 97.30% Indels: 0
DB: 18 Gaps: 0
US-10-028-384-12 (1-705) x US-10-417-375-99 (1-5404)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerThrGluLysGlnAspThrLeuLeuLysLeu 20
Db 136 AIGACTAAGTTTGGATTTTTCGATTTGCTTATGAGAGAGCAGACACATTTTGAAGCTT 195
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 196 CTCATTCTGTCAATGGCTGCTGTATTATCTCTCCACTGCTGTGTGCTGCTGCTGCA 255
Qy 41 PheGluSerValIleHisGluPheAspProThrPheAsnThrArgThrThrArgPheLeu 60
Db 256 TTTGAAAGTGTATCCAGTATGATCGTACTTTAATTATCGGACTACCAAGTTCCTG 315
Qy 61 AlaGluGluGlyPheThrLysPheHisAsnThrPheAspAspArgAlaThrThrProLeu 80
Db 316 GCTGAGGAGGGGTTTATAAATTCCTAATCTGATGATGATGATGATGATGATGATGATG 375
Qy 81 GlyArgIleIleGlyGlyThrIleThrProGlyLeuMetIleThrSerAlaAlaIleThr 100
Db 376 GGACGAAATCATTTGAGAGACAAATTTACCCAGGTTTATGATCACTCTGCTGCAATCTAC 435
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 436 CATGTACTCCATTTTTCACATCACCATCGACATTCGGAATGCTGTGTGTGCTGCTGGCC 495
Qy 121 ProLeuPheSerSerPheThrSerIleValThrThrLysLeuLeuThrLysGluLeuLysAsp 140
Db 496 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 555
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyThrIleSerArg 160
Db 556 GCAGGGGCTGGGCTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 615
Qy 161 SerValAlaGlySerThrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrThr 180
Db 616 TCTGTGGCTGGCTCTCTATGATTAATGAAGGATTTGCCATCTTTTGCATGCTACTCACTAC 675
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 676 TACATGTGGATCAAGGCGATGAAGACTGGTTTCCATCTGTGTTGGGCGAGCTAAGTGTGCCCT 735
Qy 201 AlaThrPheThrMetValSerSerTrpGlyGlyThrValPheLeuIleAsnLeuIlePro 220
Db 736 GCTTATTTCTACATGCTCTCGTCATGGGAGGTTATGTTGTTCTGTGATCACTTAATTCCT 795
Qy 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleThrValAlaThr 240
Db 796 CTCACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 855
Qy 241 CysThrValThrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db 856 TGTACTGTTTACTGCTCTGGGCACTATATCTTCTATGATGATCTCTTGTGGGTTCCTCCAG 915
Qy 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db 916 CCTGTCTCTTCTATCAGACACATGCGAGGCTTTGGGTCTTTGGGTCTCTCTCCAGATCCAT 975
Qy 281 AlaPheValAspThrLysLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
Db 976 GCCTTTGTGGATTACCTGCGCAGCAAGTTGAATCAACAATTTGAAGTCTTTTCCGG 1035
```

QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
DB 1036 AGCGTCATCTCTCGTAGGCTTTCTCTTCCACCGTGGGAGCTCTCCATCGCTGACA 1095
QY 321 GlyValIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
DB 1096 GGAATAATATCTCCCTGGAGGGGGTCTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1155
QY 341 AsnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
DB 1156 ACAACATCCCATATGCTTCTGTGTGATGATCAGCCCAACCTGGTCTCATATC 1215
QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
DB 1216 TATTTTGAACCTCGCTCTCTCTATGTTTCAGTTGGCTCTTACTGCTTATGCTAGC 1275
QY 381 AsnLeuSerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSer 400
DB 1276 AACCTGCTGTATGCCCGGATTTTATCATCATGATGATGTGACCAAGATGACTTTTCA 1335
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
DB 1336 GCTGTAAATGCTGCTAATGCTAGTGTGGCACCTGTATGTGCATTTCTCTGGCAT 1395
QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
DB 1396 GGAGTCTCCAGGTGCTGCCACATACATGAAGAACTCGGACATAAGTCTGCAGACAAG 1455
QY 441 LysSerLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
DB 1456 AAGAGCAAGACCAACAGGATTCACCTACCTATTAAGATGAAGTGGCAAGTGGGATG 1515
QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer 480
DB 1516 ATACTGCTCATGGCTTTCTTCTCATCTACCTACACCTTTCATCACTGGGTGACCACT 1575
QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
DB 1576 GAGGCTACTCTCTCCGCTCCATTTGCTATCTGCCCCGTGGTGGGATGCGACATGATC 1635
QY 501 IlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
DB 1636 ATATTTGATGATTCGAGAGCAATATTTGGCTTCGTCTATATATCTCCAGAGGATGG 1595
QY 521 LysValMetSerThrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
DB 1696 AAGGTCACTCTCTGGTGGGATTTATGGCTATCAGATTAACAGTATGGCAAAACCAAT 1755
QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
DB 1756 TTAGTGGACAAATACATGGATTAATACCCATATTTCTCAGTAGGGCAGGCANTGGCG 1815
QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
DB 1816 TCACAGAGAGAAAAGCTTATGAGATCATGAGGAGCTCGATGTGCTATGCTGGTC 1875
QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetVal 600
DB 1876 ATTTTGGAGGCTCACTGGGTATTTCTCTGATGATATCAACAGTTTCTTTGGATGGTC 1935
QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
DB 1936 CGGATTTGGAGGAGCAGATACAGGCAAAATATCAAGGAGATGACTATTATCTCCA 1995
QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
DB 1996 ACTGGGGAGTTCCTGTGGACCGTGAAGTTCTCCAGTGTCTCACTGCCTCATGTATC 2055
QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe 660
DB 2056 AAGATGTGTACTATGCTTTGGACAGGTTTACAGAGGTTTACAGAGGCAAGCGCTCTCCAGGCTTT 2115

QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
DB 2116 GACCGTGTCCGAATGCTGAGATGGGATAAAGATTTTGGCTTGATGCTCTGGAGAA 2175
QY 681 GlyTyrThrThrGluHisTyrPheValAlaGlyIleTyrLys 693
DB 2176 GCATATACCAAGAACATTTGGTGGTCCAGGATATACAAG 2214
RESULT 9
US-10-417-375-97
; Sequence 97, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandino
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001.600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 5827
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-97
Alignment Scores:
Pred. No.: 0 Length: 5827
Score: 3597.00 Matches: 686
Percent Similarity: 99.13% Conservative: 1
Best Local Similarity: 98.99% Mismatches: 6
Query Match: 97.30% Indels: 0
DB: 18 Gaps: 0
US-10-028-384-12 (1-705) x US-10-417-375-97 (1-5827)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 136 ATGACTAAGTTGGATTTTGGCATTTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT 195
QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 196 CTCAATTCGTCAATGGCTGCTGATTTATCTTCTCCACTCGTCTGTTGCTGCTCTGAGA 255
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 256 TTTGAAAGTGTATCCATGATGATTTGATCCGTACTTTTAAATATCGGACTACCGAGTTCTTG 315
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrTyrProLeu 80
DB 316 GCTGAGGAGGGGTTTATAAATCCATTAAGTGTGATGACCGAGCCTGTACCTTTG 375
QY 81 GlyArgIleLeuGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 376 GGACGAATCATTTGGAGGAACAATTTACCCAGAGTTTAAATGATCACCTCTCTCTCAATCTAC 435
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 436 CATGACTCCATTTTTCACATCAACATGACATTCGAAATGCTGTGTGTCTGTCGACC 495
QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 496 CCT 555
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
DB 556 GCAGGGGCTGGGCTTCTTGTGCTGCCATGATTTGCTGTAGTTCTCTGGATATATCTCCCGA 615
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 616 TCTGTGCTGCTCTCTATGATGATGAAGGAGGATTCCTTTTGTGATGCTACTACTACCTAC 675

181 TyrMetTrpIleYsAlaValIleThrGlySerIleCysTrpAlaAlaIleCysAlaLeu 200
 676 TACATGTGGATCAAGGAGTAAAGATGGTTCATCTGTGGGAGCTAAGTGTGCCCTT 735
 201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheIleIleAsnLeuPro 220
 736 GCTTATTCTTACATGGTCTCGTATGGGAGGTATGTGTCTCTGATCAACTTAATCTT 795
 221 LeuHisValLeuValLeuMetLeuThrGlyATGPhSerHisArgIleTyrValAlaTyr 240
 796 CTCACGTCCTCGTGTGATGCTCACAGGCGCTTCTCTCACCGGATCTATGGCCCTAC 855
 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 856 TGTACTGTCTTACTCCCTGGGACATATATCTTATGAGATCTCTTGTGGGTTCAG 915
 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 916 CCGTCTCTTCTCATCAGAGCATGGGAGCGCTTGGGCTCTTGGTCTCTGCCAGATCCAT 975
 281 AlaPheValAspTyrIleLeuArgSerLysLeuAspProGlnGlnPheGluValLeuPheArg 300
 976 GCGTTTGTGATACCTCCGACGAGAGTGAATCCACAAATTTGAAGTCTCTTTTCCGG 1035
 301 SerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuMetLeuThr 320
 1036 AGCGTCATCTCTCGTAGGCTTGTCTCTCACCGTGGAGCTCTCTCATGTGTGACA 1095
 321 GlyIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 1096 GGAATAATATCTCTCGAGGGGCGTCTCTCTCTGCTGTGATCCCTCTTATGCTAAG 1155
 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrTrpSerSerTyr 360
 1156 AACACATCCCATCATTTGCTGTCTGTGAGCATCAGCCACACCTGTGTCCTCATAC 1215
 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 1216 TATTTTGNACCTCGAGCTCTCGTCTTCATGTTTCCAGTTGGCTCTTATGCTGTAGC 1275
 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 1276 AACCTGTCTGATGCGCGGATTTTATCATCATGTATGCTGTGACCAAGCATGTACTTTCA 1335
 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 1336 GCTGTATGCTGCTTATGCTAGTGTGGACCTGTTATGTGCTATCTCTCTGGCAT 1395
 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
 1396 GGAGTCTCCAGGTGCTGTCCACATACATGAAGAATCTGGACATAGTGTCCAGACAA 1455
 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
 1456 AAGAGCAGAGCAACAGGATTCACCTACCTATTAAGAAAGAGTGGCAAGTGGGATG 1515
 461 IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer 480
 1516 ATACTGTGTCATGGCTTCTTCTCTCATCACCCTACACCTTTTCACTACCTGGGTGACCA 1575
 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyValAspGlySerArgIle 500
 1576 GAGGCTACTCTCTCTCGTCCATTTGATCTATCTGCGGTGGTGGGATGGAGTAGGATC 1635
 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 1636 ATATTGTATGATCTCCAGAGCATATATTATGCTTCTGTCATATATCTCCAGAGATCG 1695
 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
 1696 AAGGTCATGCTCTGTGGGATATGGCTATCAGATTACAGCTATGCAACACCCGACAA 1755
 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560

1756 TTAGTGGCAATAACACATGGAATAATACCATATTTCTCGAGTAGGCGAGCAATGGCG 1815
 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 1816 TCCACAGAGGAAAGCCATGATGATCATGAGGAGCTCGATGTCAAGTATGTCTGGTC 1875
 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrpMetVal 600
 1876 ATTTTGGAGGCTCCTCCTGATATCTCTGATATCAACAGTTTCTTTGGATGGTC 1935
 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
 1936 CGGATTTGGAGGAGCACAGATACAGGCAACATATCAAGGAGATGACTATTATCTCCA 1995
 621 ThrGlyGluPheArgValAspArgGluSerProValLeuLeuAsnCysLeuMetTyr 640
 1996 ACTGGGAGTTCCTGTCGACCGTGAAGGTTCTCCAGTGTCTGCTCACTCCCTCATGTAC 2055
 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 2056 AAGATGTGTACTATCTGCTTTGGACAGTTTACACAGAGCAAGCGTCTCTCAGGCTTT 2115
 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 2116 GACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCGGAGAA 2175
 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLys 693
 2176 GCATATACACAGACATTTGCTGTCAGGATATACAAG 2214

RESULT 10

US-10-425-115-150745
 ; Sequence 150745, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 150745
 ; LENGTH: 2839
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: M874577_69000C.1
 ; US-10-425-115-150745

Alignment Scores:
 Pred. No.: 1,178-222 Length: 2839
 Score: 2258.00 Matches: 420
 Percent Similarity: 75.91% Conservative: 122
 Best Local Similarity: 58.82% Mismatches: 142
 Query Match: 61.08% Indels: 30
 DB: 18 Gaps: 7
 US-10-028-384-12 (1-705) x US-10-425-115-150745 (1-2839)

QY 7 LeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeuLeuLeuLeuSerMetAla 26
 DB 254 CTGCGCTCAAGACCAAGCAGCAGGAGCTCTGTCTCCGCTCGGCGCTGGCGCTCATC 313
 QY 27 AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis 46
 DB 314 TACGTGCTGCTCTCGCGCTCGCTCTCTCTCCGTCTCCGCTACGAGTCCATGATCCAC 373
 QY 47 GluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAlaGluGluGlyPheTyr 66

Db 374 GAGTTCGACCCCTTACTTCAACTACCGCACCGACCGTCTCTCTCAACGAGACCGCTTCAGC 433
Qy LysPheHisAsnTrpPheAspArgAlaTrpTyrProLeuGlyArgIleGlyGly 86
Db 434 GAGTTCGACCGTCTGACCTCGAGAGCTGGTACCGCTCGCGCGCGTGTGTCGGGGG 493
Qy ThrIleTyrProGlyLeuMetIleThrSerAlaIleTyrHisValLeuHisPhePhe 106
Db 494 AGCTTCTACCCGGGTCTGATGTGACCGCGCGGTCTCCACCGCCCTCTCCCGCGCTC 553
Qy HisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPhe 126
Db 554 TACCTCACCGTCCACATCGGTGAGGTCTGCGTCTCCACCGCGCGCTCTCTCGCTGCAAC 613
Qy ThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyIleLeu 146
Db 614 ACCAGCTCTGCTGCTACCGTTCGCGCGCGAGATGGGACTCCGGCGCGGCTGCTC 673
Qy 147 AlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAlaGlySerTyr 166
Db 674 GCGCTGCGCTGATCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 733
Qy 167 AspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTyrIleLysAla 196
Db 734 GACACGAGGCGTCCCATCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793
Qy 187 ValLysThrGlySerIleCysTrpAlaLysCysAlaLeuAlaTyrPheTyrMetVal 206
Db 794 GTTAACACTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 853
Qy 207 SerSerTrpGlyTyrValPheLeuIleAsnLeuIleProLeuHisValLeuValLeu 226
Db 854 TCGGATGGGAGGTAGTGTTCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913
Qy 227 MetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeu 246
Db 914 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 973
Qy 247 GlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerSerGlu 266
Db 974 GGGATGTCTTGAATGCAATTCGTTTGTGATTCAGCATGTCAGATGCGAGTCCGGAG 1033
Qy 267 HisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeu 286
Db 1034 CACATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1093
Qy 287 ArgSerLysLeuAsnProGlnPheGluValLeuPheArgSerValIleSerLeuVal 306
Db 1094 AAATATCTGCTAAATGATGCCAGA-----CTATTCAAGTCATTCTCTGCAATTACC 1144
Qy 307 GlyPheValLeuThrValGlyAlaLeuMet-----LeuThrGlyLys 322
Db 1145 CTCACATGTGATTACTATTGGCACCCCTGCTCTGGACTGCTGCTGCTGCTGCTGCTGCT 1204
Qy 323 IleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLysAsnAsn 342
Db 1205 ATCTCCCTTGGACAGACGGTTTACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1264
Qy 343 IleProIleAlaSerValSerGluHisGlnProThrThrThrThrThrThrThrThrThr 362
Db 1265 ATACCAATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1324
Qy 363 AspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeu 382
Db 1325 GACTTCCACATCT 1384
Qy 383 SerAspAlaArgIlePheIleLeuMetTyrGlyValThrSerMetTyrPheSerAlaVal 402
Db 1385 TCAGATGCCACAATATTATTGTTATGATGCGCTCACAAGATGATGATGATGATGATGATGAT 1444
Qy 403 MetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGlyVal 422

1445 ATGTCGCGTTAATCTTTGTCGACGACGACGCGTTCCTTATTTAGTCTATTGCTGCA 1504
423 SerGlnValLeuSerThrTyrMetLysAsnLeu---AspIleSerArgProAspLysLys 441
1505 TCTGCTCAATA-----AAAAACCTAACCACTTTGATCAGGACCAAGAGCAGA 1552
442 Ser-----LysLysGlnGln 446
1553 AGTCCACGACTACTTCTGGGAAAAACAAGTCCAAAGTCAGCTCCAAAGGGTGCAGTT 1612
447 AspSerThrTyrProIleLysIleGluValAlaSerGlyMetIleLeuValMetAlaPhe 466
1613 GATCAATCTTTACCATTTCCACACAATGCTGCTATTGGTCTCTCTCTGCGTCTTTAC 1672
467 PheLeuIleThrTyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerPro 486
1673 TTGCTCAGTAGGTATGCCATACACTGCACCTTGGGTGCATCTGAGGCTTACTCTCTCCA 1732
487 SerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePheAspPheArg 506
1733 TCTATAGTTTGGCGCGACGAGGTTCATAATGGTGGAGAGTCACTCTTGTATGATTATCGT 1792
507 GluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrp 526
1793 CAGGATACACTATTGGCTTCGTGACACACGCTCTCTGATGCTAAGATTATGCTGCTGG 1852
527 AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThr 546
1853 GACTATGATACCAAAATACAGCTATGGGTAAACAGAACTGTATTGTTGATACCAATACA 1912
547 TrpAsnAsnThrHisIleSerArgValGlyGluAlaMetAlaSerThrGluGluLysAla 566
1913 TGGATTAACACACATAGTCTACAGTTGACGAGCTATGCTATCTTATGAGATGAGCA 1972
567 TyrGluLeuMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThr 586
1973 TATGAGATCATGCACTGCTGATGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2032
587 GlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThr 606
2033 CGATATCTTCAGATGATATCAACAGTTCTTATGATGCTGCTGCTGCTGCTGCTGCTGCT 2089
607 AspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGlyGluPheArgVal 626
2090 ---GTTTTCTCTGTAATCAAAAGAGCCAGATTACCTGTT---AATGGGAGTATCGTATT 2143
627 AspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArg 646
2144 CACAAGGGGGCAGCACCCCAAAATGCTGAATGCTGCTAATGTACAAAGCTTTGTACTACCA 2203
647 PheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArgValArgAsnAla 666
2204 TTTGGAGAACTTACACAGAAATATGAAAAACCTCCAGGGTATGCTGCTGCTGAGGAGTG 2263
667 GluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrThrGluHis 686
2264 GAGATTGGCAACAAGACATAAGCTTGAATCTTGGAGGAGGCAATTACACTTCAAC 2323
687 TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
2324 TGGATAGTGGCATATACAAGGTGAAACCCCAAGAACACAGG 2365

RESULT 11
US-10-425-114-14408
Sequence 14408, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabasta, Jack E
APPLICANT: Cao, Yongwei

Db 2000 TATCTTCAGATGATATCAACAGTCTTATGGATGGTACGTAATGGTGGAGCA----- 2053
 QY 608 ThrGlyLysHslelleGluAsnAspTyrThrProThrGlyGluPheArgValAsp 627
 Db 2054 GTTTTTCCTGTAATCAAGAGCCAGATTACCTTGT---AATGGGAGTATCGTATTGAC 2110
 QY 628 ArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPhe 647
 Db 2111 AAGGGGGGAGCCCAAAATGCTGAACTGCTTAATGTAACAGCTTTGTTTACTACCGATT 2170
 QY 648 GlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGlu 667
 Db 2171 GGAGAACTTACCACAGATATGGAACCTCCAGGGTATGATCGTACGAGAGTGGAG 2230
 QY 668 IleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHisTrp 687
 Db 2231 ATGGCAACAGACATAAAGCTTGATACTTTGGAGAGGCAATTACACTTCAAACTGG 2290
 QY 688 LeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2291 ATAGTGGCATATACAGGTGAACCCCAAGAACACAGG 2329

RESULT 12

US-10-028-384-1
 ; Sequence 1, Application US/10028384
 ; Publication No. US20030148285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COMPATIGENE INC.
 ; APPLICANT: PERREAU, Claude
 ; APPLICANT: MCBRIDE, Kevin
 ; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
 ; FILE REFERENCE: 5600-74
 ; CURRENT APPLICATION NUMBER: US/10/028,384
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 2481
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2481)
 ; OTHER INFORMATION:
 US-10-028-384-1

Alignment Scores:
 Pred. No.: 1,3e-221 Length: 2481
 Score: 2247.00 Matches: 427
 Percent Similarity: 73.99% Conservative: 102
 Best Local Similarity: 59.72% Mismatches: 150
 Query Match: 60.78% Indels: 36
 DB: 15 Gaps: 7

US-10-028-384-12 (1-705) x US-10-028-384-1 (1-2481)

QY 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaValLeuSerPheSerThrArgLeu 35
 Db 208 TCGCTTCCTCTCCATCCATCTCTCTCGCTGGCTGGCTTCAGCTCGCGCCCT 267
 QY 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
 Db 268 TTCGGCGTCATCGCTTCGAAAGCATCATCCACGAGTTCGACCCGCTGTTAACTATAGA 327
 QY 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspAspArg 75
 Db 328 TCACACATCATCTTCGACCTCATGGGTCTATGAATTTTAAATTTGGTTTGATGAAGA 387
 QY 76 AlaTyrProLeuGlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThr 95
 Db 388 GCATGTATCCACTAGGAAGATAGTGGTGTCTGTTTACCCAGGTTGATGATAACC 447
 QY 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115

Db 448 GCTGGCCTTATCTTATGGATTTTAAATACATTAACATGTTCCACATAGAGACGTA 507
 QY 116 CysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeuThr 135
 Db 508 TGTGTGTTCTTGTGCAACAACTTTTAGCGGCTTTACATCTATATCTACTTCTCTGCTTACA 567
 QY 136 LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
 Db 568 AGAAGACTTTTGGAAACCAAGGAGGAGGACTTTTAGCTGCTGTGTTTATTGCTATTGTACCA 627
 QY 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 628 GGCTACATATCTCGGTGAGTAGCTGATCTCTTTGATAATCAAGGCACTTGTATTTTGA 687
 QY 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
 Db 688 CTTGAGTTTCATCTACTATTATGGTAAATCTGTAAACCTGGTCACTTTTGTGACA 747
 QY 196 AlalysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
 Db 748 ATGTGCTGCTCTTATCTCTATATGCTCTGCTTGGGTGCTTATGTATTTATC 807
 QY 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 Db 808 ATCAATCTTATCTCCACTGCAATGTTTGTGTTTACTGATGAGAGATACAGCAAAAGA 867
 QY 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 Db 868 GTCTACATAGCATATAGCACTTCTCATATGTTGGTTTAAATATTAATCAATGAGATACCT 927
 QY 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 Db 928 TTTGTGGATTTCCAGCCATCAAGAACAGTGAACATGACATGACGCTGCAGTGTCTTTGCA 987
 QY 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnIlePhe 295
 Db 988 TTGCTGCAAGCTTATGCTTCTTGTGAGTATCTGAGAGCCGATTAACAAACAAAGAGTTC 1047
 QY 296 GluValLeuPheArgSerValIleSerLeu---ValGlyPheValLeuLeuThrValGly 314
 Db 1048 CAGACCTTTTCTTTTGGGTGATATCACTAGCTGCGAGTGTCTGTCTTCTTAGTGTCT--- 1104
 QY 315 AlaLeuLeuMetLeuThrGlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeu 334
 Db 1105 ATCTATTGACTTATACAGGTTTACATTTGCACCATGAGTGGCAGGTTTATTCTATTGCTG 1164
 QY 335 AspProSerTyrAlaLysAsnAsnIleProIleAlaSerValSerGluHisGlnPro 354
 Db 1165 GATACCTGGTATGCAAAAATACATCTCCAATTTATGCAATGATGCTGCTGAGCATCACT 1224
 QY 355 ThrThrTrpSerSerTyrPheAspLeuGlnLeuValPheMetPheProValGly 374
 Db 1225 ACGACTTGGGTGTCTTCTTCTTGTATCTACATATCTTGTATGATACCTTCCAGCAGGC 1284
 QY 375 LeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyVal 394
 Db 1285 CTTTGGTTCGATCAAAAATATCAAGCATGAAGAGATTTTGTCTCTATATGCAATC 1344
 QY 395 ThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
 Db 1345 AGTGTGTCTACTTTGTGAGTGTGATGCTGAGTGTGATGCTGCTGCTGCTGCTGCTGCTG 1404
 QY 415 SerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAsp 434
 Db 1405 TGTATGCTGCTGCAATGCTTTTCAATGTTTGTGAGCACTATTGTTGGGAT---GAC 1461
 QY 435 IleSerArgPro----- 438
 Db 1462 ATGAAAAGGGAATAATCCACTGTGGAGACAGCAGTATGAGGATGACAAAAGAACCA 1521
 QY 439 -----AspLysLysSerLysLysGlnAspSerThrTyrProIleLysIle 454


```

1522 GGAATTTGTATGATAGGAGGTAAGTGAAGAAACATGCACTGAACAGGAAAAAACT 1581
Qy 455 Glu-----ValAlaSerGlyMetLeuValMetAlaPhePheLeu 468
Db 1582 GAAGAGGATTAGGCCCTTAATAAAGCAATGTCACCATTTGATGCTGATGCTATTG 1641
Qy 469 IleThrTrpPheHisSerThrTrpValThrSerGluAlaThrSerSerProSerIle 488
Db 1642 ATGATGTTGCTGCTCACTACTGCTGGTCACAGCAATGCTTCTTCTGATGCTGATG 1701
Qy 489 ValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAla 508
Db 1702 GTCTGGCCCTCATACATCATGATGTCAGCAGCAATATCTTAGATGATTTAGAGAAGCT 1761
Qy 509 TyrTrpTrpLeuArgHisAsnThrProGluAspAlaValMetSerTrpTrpAspTyr 528
Db 1762 TACTTTGGCTTAGGCAAAATACAGATGAACATGACGAGTAATGCTCTGTTGGGATTA 1821
Qy 529 GlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnThrTrpAsn 548
Db 1822 GCCTATCAGATAGCTGGAATGGCTAATAGAACTACGTTGGTGGATAATAACACCTGGA 1881
Qy 549 AsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGlu 568
Db 1982 AACAGCACATAGCAGCTGGTGGAAAGCTATGCTCTTAATGAACAGCAGCCTATAA 1941
Qy 569 IleMetArgGluLeuLeuAspValSerTyrValLeuValIlePheGlyGlyThrGlyTyr 588
Db 1942 ATCATGAGGACTCTAGATGATGATATGTTTGGTATATTTTGGAGGGGTTATTGGCTAT 2001
Qy 589 SerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThr 608
Db 2002 TCTGGTATGATATCAACAATTTCTCTGATGGTTAGATAGCT-----GAA 2049
Qy 609 GlyLysHis-----IleLysGluAsnAspTyrTyrThrProThrGlyGluPheArg 625
Db 2050 GGAGAACATCCCAAGACATTCGGAAAGTGAATATTTACCCACAGGAGAAATTCGGT 2109
Qy 626 ValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyr 645
Db 2110 GTACACAAAGCAGGATCCCTACTTTGTTGAATGCTTATGATATAAATGTACACTAC 2169
Qy 646 ArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArgValArgAsn 665
Db 2170 AGATTGGAGAAATGCAGCTGGATTTTCGTACACCCCGCAGGTTTTCACCGAACAGTAA 2229
Qy 666 AlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGluGlyTyrThrThrGlu 685
Db 2230 GCTGAGATTGGAATAGGACATTAATCAACATTTGGAAGAGAGCCTTACATACAGA 2289
Qy 686 HistTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
Db 2290 CACTGGCTTGTAGGATATATAAGTAAAGCAGCCTGATAACAGG 2334

```

RESULT 13
US-10-028-384-3
; Sequence 3, Application US/10028384
; Publication No. US20030146285A1
; GENERAL INFORMATION:
; APPLICANT: COMPATIGENE INC.
; APPLICANT: PERREAU, Claude
; APPLICANT: MCBRIDE, Kevin
; TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy
; FILE REFERENCE: 5600-74
; CURRENT APPLICATION NUMBER: US/10/028,384
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Mus musculus
; PUBLICATION INFORMATION:

```

; DATABASE ACCESSION NUMBER: AK018758
; DATABASE ENTRY DATE: 2001-07-05
; RELEVANT RESIDUES: (1)..(2469)
US-10-028-384-3

Alignment Scores:
Pred. No.: 1,628-220 Length: 2710
Score: 2237.00 Matches: 424
Percent Similarity: 73.57% Conservative: 102
Best Local Similarity: 59.30% Mismatches: 153
Query Match: 60.51% Indels: 36
DB: 15 Gaps: 7

US-10-028-384-12 (1-705) x US-10-028-384-3 (1-2710)

Qy 16 ThrLeuLeuLysLeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu 35
Db 275 TGTGTGCTCTCTTCCATCTCTTCTGCGCTGCTGCGCGCTTCAGCTCGCGCTC 334
Qy 36 PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg 55
Db 335 TTGCGCGCTCATCGCTTCGAGAGCATCATCCAGGATTCGACCGCTGTTTAACTAGA 394
Qy 56 ThrThrArgPheLeuAlaGluGlyPheTyrLysPheHisAsnTrpPheAspAspArg 75
Db 395 TCACACATCATCTTCATCTCTCATGATCTCATGATTTCTAAATGTTGTTGAAAGA 454
Qy 76 AlaTyrProLeuGlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThr 95
Db 455 GCATGTGATCCCACTGGGAGAAATAGTGGTGGCACCCTTTACCAGGTTGATGATAACA 514
Qy 96 SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal 115
Db 515 GTGCGCTTATTCATGATGATTTAAATACATGAGCAATTAACAGTTTCATTAAGAGATG 574
Qy 116 CysValPheLeuAlaProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThr 135
Db 575 TGTGTATTCTTGTGACCAACTTTTAGCGGCTTACATCCATATCTACGTTCTCTGCTAACT 634
Qy 136 LysGluLeuLysAspAlaGlyValAlaGlyLeuLeuAlaAlaMetIleAlaValPro 155
Db 635 AGAGAACTGTGACCAAGAGCAGGACTTCTAGCTGCTGCTTCATCTGATGTCGATACCA 694
Qy 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
Db 695 GGTACATATCTCGGTGATGCGGGATCTTGTATTAATGAAGGATTCGATTTTTCG 754
Qy 176 MetLeuLeuThrTyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAla 195
Db 755 CTTCAGTTCATCTACTTATGGTAAAGTCTGTGAAGCCGCTGTGTTCTGTGACA 814
Qy 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeu 215
Db 815 ATGTGCTGCTGCTTGTATTTTACATGCTCTCTGCTGCGGAGGTTATGTTGTCATC 874
Qy 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
Db 875 ATCAACCTCAUCCCTCTCCATGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 934
Qy 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
Db 935 GTCTACATAGCATATAGCATTGTTGATCTTGTGGGTTTAAATATTATTCATCGCAGATACCT 994
Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
Db 995 TTTTGGGATTTACGCCAATCAGAACAGGACATGGCAGCTGCGAGTGTCTTTCG 1054
Qy 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
Db 1055 CTGCTGCAAGCTTACGCTTTTTCAGTATCTGAGAGCCGCTTGACAAACAGGAGTTTC 1114
Qy 296 GluValLeuPheArgSerValIleSerLeu--ValGlyPheValLeuLeuThrValGly 314

```

Db 1115 CAGACCCCTTTCTTTTGGGTGCTCTACTAGCTGAGCGGTGTGTTCCTTAGTGTCT-- 1171
Qy 315 AlaLeuLeuMetLeuThrGlyIleSerProThrThrGlyArgPheThrSerLeu 334
Db 1172 ATCTATCTGACATACACAGGTATATTCACCATGGAGTGGAGGTATTTATTCATG 1231
Qy 335 AppProSerThrAlaLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnPro 354
Db 1232 GATAGTGGGTATGCAAAATACACATTCCTCAATATTGTCATCAGTGTCTGAACATCAGCCT 1291
Qy 355 ThrThrTrpSerSerThrPheAspLeuGlnLeuValPheMetPheProValGly 374
Db 1292 AGCATATGGGTCTCTTTCTTTGATCTACATATCTTTGATGTACCTTCCCGAGGC 1351
Qy 375 LeuThrThrCysPheSerAsnLeuSerAspAlaArgIlePheIleMetThrGlyVal 394
Db 1352 CTATGGTCTGTCATCAAAATATCAACCATGAAAGAGTATTTGCTGCTGTATGGATC 1411
Qy 395 ThrSerMetThrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
Db 1412 AGTGTGTGTTACTTTGCGGAGTATGTCGGGTGATGTCGACTGACCCCGGTCTG 1471
Qy 415 SerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrThrMetIleAsnLeuAsp 434
Db 1472 TGCATGCTGTCGGCATCGCTCTCTCAATGTTTTCAGCAGTATTTGGGGAT--GAC 1528
Qy 435 IleSerArg-----ProAspLysIleSerLysGlnAspSerThrThrPro 451
Db 1529 ATGAAAGGAAACCCCTGTCGAGCAGCAGTATGAGTATGAGGATGACAAAGAACCC 1588
Qy 452 IleLysIleGluValAlaSerGlyMetIle----- 461
Db 1589 GGAACTTGTATGACAGGCGAGTAAAGTGAAGGATGTGACAGAGCAAGAGAAACCT 1648
Qy 462 -----LeuValMetAlaPhePheLeu 468
Db 1649 GAAGAGGCTGGGCCCCCAACATCAAAAGCATGTCAGCATGTCATGCTCTCTG 1708
Qy 469 IleThrThrThrPheHisSerThrTrpValThrSerGluAlaThrSerProSerIle 488
Db 1709 ATGATGTCGGGTCTCACTGACGTGGGTCAACAGCAACGCTACTCCAGTCAAGTGT 1768
Qy 489 ValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAla 508
Db 1769 GTCTTGTCTCTCAATCATGATGTTACAGGATATATTTAGATGATTTTGAAGAGCG 1828
Qy 509 TyrThrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspThr 528
Db 1829 TACTTTTGGCTGAGACAAACACCGATGAACACGCGCGGTCTATGCTGCTGGGACTAC 1888
Qy 529 GlyThrGlnIleThrAlaValAlaAsnArgThrIleLeuValAlaAsnAsnThrTrpAsn 548
Db 1889 GGCTATCATGTTGTCGATGCGCAACAGGACCACTCTGTGGGATATCAACACCTGGAAC 1948
Qy 549 AsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaThrGlu 568
Db 1949 AACAGCCATCGCATCGCTGGGAAAGCTATGCTTCCAATGAACGCGCGCTATAAA 2008
Qy 569 IleMetArgGluLeuAspValSerThrValLeuValIlePheGlyGlyLeuThrGlyThr 588
Db 2009 ATCATGAGTCCCTTGTATGATATGTTGTGTTATTTTCGAGAGGATGATGGCTAT 2068
Qy 589 SerSerAspAspIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThr 608
Db 2069 TCCGGGAGCATATCAACAGTTCCTCTGGATGCTCAGGATAGCT-----GAA 2116
Qy 609 GlyLysHis-----IleLysGluAsnAspThrThrProThrGlyGluPheArg 625
Db 2117 GGGAGCATCCCAAGACATCGGGAAGGTGACTATTTTCCACGAGGAGGATGCCGA 2176
Qy 626 ValAspArgGluGlySerProValLeuLeuAsnCysLeuMetThrLysMetCysThrThr 645
Db 2177 GTAGACAAAGCTGGGTCTCTACTCTGTAACTGCCTTATGATATAAAAGTCACTACTAC 2236

Qy 646 ArgPheGlyGlnValThrThrGluAlaLysArgProGlyPheAspArgValArgAsn 665
Db 2237 AGATTTGGAGAAATGACAGCTAGATTTTCGCACTCCCCAGGCTTTCACCGAACAGTAAT 2296
Qy 666 AlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyThrThrThrGlu 685
Db 2297 GCTGAGATGGAATAAAGACATTAATTCAGCATTTGGAGGAAGCTTTTACATCAGAG 2356
Qy 686 HisTrpLeuValArgIleThrLysValLysAspLeuAspAsnArg 700
Db 2357 CACTGGCTTGTACGATATATATAAGTGAAGACACTGACACACAGG 2401

RESULT 14
US-10-437-963-99904
; Sequence 99904, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 99904
; LENGTH: 2779
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_97672C.1
US-10-437-963-99904

Alignment Scores:
Pred. No.: 1,69e-220 Length: 2779
Score: 2237.00 Matches: 419
Percent Similarity: 75.49% Conservative: 120
Best Local Similarity: 58.68% Mismatches: 145
Query Match: 60.51% Indels: 30
DB: 17 Gaps: 7

US-10-028-384-12 (1-705) x US-10-437-963-99904 (1-2779)

Qy 7 LeuArgLeuSerThrGluLysGlnAspThrLeuLysLeuLeuLeuLeuSerMetAla 26
Db 172 CTGGCTGGAAGACGAAAGCAGCAGAGAGCTCTCTCGGCTTCGGGCTCGGCGCTGCGCTCATC 231
Qy 27 AlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerValIleHis 46
Db 232 TACGTGCTGGGTTCGTGCTGCTGCTCTCTCGGCTTCGGCTTCAGTCCATGATCCAC 291
Qy 47 GluPheSerProThrPheAsnThrArgThrArgPheLeuAlaGluGluGlyPheThr 66
Db 292 GAGTTCGATCCCTACTTCAACTACCGCACCGCTCTCTCTCCGACCATGATTCAGC 351
Qy 67 LysPheHisAsnTrpPheAspArgAlaTrpThrProLeuGlyArgIleGlyGly 86
Db 352 GAGTTCGAACTGGTTCGATTTCCAGAGAGCTGCTACCGCTCGGCGCGCTGCTGCGCGGC 411
Qy 87 ThrIleThrProGlyLeuMetIleThrSerAlaAlaIleThrHisValLeuHisPhePhe 106
Db 412 ACCCTTCTCCGGGCTCTATGCTACCGCGCGCTGCTCCACCGCTCTCTCGCGCGCTC 471
Qy 107 HisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSerPhe 126
Db 472 TCCCTCGCGCTCCACATCCGCGAGGTCTGGGTCTCTCACCCTCCCTCTTTCGCGCGCAAC 531

| | | | | | | | |
|----|------|--|------|--|------|---|------|
| QY | 127 | ThrSerIleValThrTyrLeuLeuThryesleuLeuysAspAlaGlyLeuLeu | 146 | QY | 467 | PheLeuIleThrTyrThrPheHisSerThrTrpValThrSerGluAlaTyrSerSerPro | 486 |
| DB | 532 | ACCAGCGTCGTCAGCCCTCGCGCGGAGATCTGGAGACTCGCGCGGCGCTCGTC | 591 | DB | 1591 | TTGCTTAGTATAGTATGCTGTACATGTCACCTTGGGTGACATCTGAAGCTTACTCTCTCT | 1650 |
| QY | 147 | AlaAlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAlaGlySerTyr | 166 | QY | 487 | SerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArg | 506 |
| DB | 592 | GCCGCGCGCTCATCGCGCTGTCGCGCGGCTACATCTCCGCGCTCGCGCGGCTCTAC | 651 | DB | 1651 | TCATTTGTTCTGGCGCAAGGGGTCTAATAGGAGCGAGGGGTCTATTATTGATGATTTCGT | 1710 |
| QY | 167 | AspAsnGluGlyIleAlaIlePheCysMetLeuLeuThryTyrMetTrpIleLysAla | 186 | QY | 507 | GluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrp | 526 |
| DB | 652 | GACACAGAGGGGTCGCCATCTTCGCGCTGCTGCTCACCTCTACTCTTCGTCGCGGC | 711 | DB | 1711 | GAAGCGTATTATGCTTCTGTCAGAACACTCTCTCTGATGCAAGATTATGTCATGTGG | 1770 |
| QY | 187 | ValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPheTyrMetVal | 206 | QY | 527 | AspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnThr | 546 |
| DB | 712 | GTCAACACGGGATCTCTCGCATGCTGCTCGCTCGCGCTTCGCGCTACTTCTCATGTGC | 771 | DB | 1771 | GATTATGATACCAATCACAGCTATGGGGAACAAGACTGTTATTGTTGATACCAACA | 1830 |
| QY | 207 | SerSerTrpGlyTyrValPheLeuIleAsnLeuIleProLeuHisValLeuValLeu | 226 | QY | 547 | TrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAla | 566 |
| DB | 772 | TCCGCGTGGGAGGCTATGCTTCTCATCATCAATCTGTTCGCGCTCTACGTCGTCGTC | 831 | DB | 1831 | TGGAATAATACACATATACGCTACAGTTCGGCGGTGCAATGTCATCTCTATGAAGATGA | 1890 |
| QY | 227 | MetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrValTyrCysLeu | 246 | QY | 567 | TyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThr | 586 |
| DB | 832 | CTCGTCACGGGAGGTAATCTCGCAGAGGCTCTACGTCGCGCTCAAAATCCACGATGTC | 891 | DB | 1891 | TATGAGATAATGCAGTCACTGATGCTGAATATTATGCTGTCGTTGCTATTGGTGGTACT | 1950 |
| QY | 247 | GlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeuSerSerGlu | 266 | QY | 587 | GlyTyrSerSerAspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThr | 606 |
| DB | 892 | GGAAATGCTACTTTCGATGCGATCGGTTCTGTTGGTTCCAAACAGCTCCAGTCTGGGAG | 951 | DB | 1951 | GGTTATTTCTCAGATGACATTAACAAGTCTTATGATGGTGGCGATTGGTGGTGA | 2007 |
| QY | 267 | HisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeu | 286 | QY | 607 | AspThrGlyLysHisIleLysGluAsnAspTyrTyrTrpProThrGlyGluPheArgVal | 626 |
| DB | 952 | CACATGCGCGCTATGGAGTATCTTCTCTTTGACAGTTTCTTTCTTCTGACTGGGTG | 1011 | DB | 2008 | ---GTTTTCCTGTAATCAAGAGCGCTGATTACCTTGT---AATGGGAGTATCGTGC | 2061 |
| QY | 287 | ArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuVal | 306 | QY | 627 | AspArgGluGlySerProValLeuLeuLeuLeuLeuMetTyrLysMetCysTyrTyrArg | 646 |
| DB | 1012 | AAATACCTCTGAATGATGCCAAA-----CTATTCAAGTCTTCTCAGAAATCACC | 1062 | DB | 2062 | GACAGGGGCGACACCAAAATGTTGAATGCTTAAATGTAACAGCTTTGTTATTACCGT | 2121 |
| QY | 307 | GlyPheValLeuLeuThrValGlyAlaLeuLeuMet-----LeuThrGlyLys | 322 | QY | 647 | PheGlyGluValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAla | 666 |
| DB | 1063 | CTGACATGTGTGATAACTGTGGCACCTTAGCTTGGAAATGTTGACCGCGCTCAGGTATC | 1122 | DB | 2122 | TTTGGAGAACTGACACAGANATGGAACCTCCAGGATACGATCAGTACGAGGAGTG | 2181 |
| QY | 323 | IleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaIysAsnAsn | 342 | QY | 667 | GluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrTrpThrThrGluHis | 686 |
| DB | 1123 | ATCTCCCTTGGACAGCGGGTTTACTCCCTGCTGATCCAACTATGCAAGAGGACCAT | 1182 | DB | 2182 | GAGATTGGCAACAACAGACATTAAGCTTGAGTACTTGGAGGAGGCATTCCACACTTCGAAC | 2241 |
| QY | 343 | IleProIleIleAlaSerValSerGluHisGlnProThrTrpSerSerTyrTyrPhe | 362 | QY | 687 | TrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg | 700 |
| DB | 1183 | ATACCAATCATTCATCATGATCTGAGCATCAACCAAGCGTGGTCTCGTTTCATGTTT | 1242 | DB | 2242 | TGGATAGTGGCATATACAAAGTCAAAACCCCGAGAGATAGG | 2283 |
| QY | 363 | AspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeu | 382 | RESULT 15 | | | |
| DB | 1243 | GAITTCACATCTCTTTCTTCTTCTTCCAGCAGCGCTCTACTTCTGTTTCAAGCGCTG | 1302 | US-10-028-384-7 | | | |
| QY | 383 | SerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSerAlaVal | 402 | Sequence 7, Application US/10028384 | | | |
| DB | 1303 | TCAGATGCCACAATATTATAGTATGATGGCTCACGATGATGATGCTGCTGCTG | 1362 | Publication No. US20030148285A1 | | | |
| QY | 403 | MetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGlyVal | 422 | GENERAL INFORMATION: | | | |
| DB | 1363 | ATGTTTCGGTTAATCTTGTGTCAGCGCCAGCTGTTTGTGCTTATTAGTCCCATTCGAC | 1422 | APPLICANT: COMPATIGENE INC. | | | |
| QY | 423 | SerGlnValLeuSerThrTyrMetLysAsnLeu---AspIleSerArgProAspLysLys | 441 | APPLICANT: PERREAU, Claude | | | |
| DB | 1423 | TCAGCTACAATA-----AAAACTGACTACTTGTGTTGCGCAAAAGCAAA | 1470 | APPLICANT: MCBRIDE, Kevin | | | |
| QY | 442 | Ser-----LysLysGlnGln | 446 | TITLE OF INVENTION: Mammalian SIMP, gene sequence and uses thereof in cancer therapy | | | |
| DB | 1471 | AGTCCACAAACTGCTCTGGAATAATCATCTGCTCAAGAGCAGCTCAAAAGGGGTGCGATT | 1530 | FILE REFERENCE: 5600-74 | | | |
| QY | 447 | AspSerThrTyrProIleLysIleGluValAlaSerGlyMetIleLeuValMetAlaPhe | 466 | CURRENT APPLICATION NUMBER: US/10/028,384 | | | |
| DB | 1531 | GATCAATCTTGGCTTCCAGCAGATGATGGCCATTGCTTACTTCTGGGTGCTTCTAC | 1590 | CURRENT FILING DATE: 2001-12-20 | | | |
| | | | | NUMBER OF SEQ ID NOS: 13 | | | |
| | | | | SOFTWARE: Patent in version 3.1 | | | |
| | | | | SEQ ID NO 7 | | | |
| | | | | LENGTH: 2417 | | | |
| | | | | TYPE: DNA | | | |
| | | | | ORGANISM: Drosophila melanogaster | | | |
| | | | | PUBLICATION INFORMATION: | | | |
| | | | | DATABASE ACCESSION NUMBER: AF132552 | | | |
| | | | | DATABASE ENTRY DATE: 1999-04-27 | | | |
| | | | | RELEVANT RESIDUES: (1)..(2417) | | | |
| | | | | US-10-028-384-7 | | | |
| | | | | Alignment Scores: | | | |

Pred. No.: 1-95e-214 Length: 2417
 Score: 2177.50 Matches: 415
 Percent Similarity: 72.92% Conservative: 110
 Best Local Similarity: 57.64% Mismatches: 144
 Query Match: 58.90% Indels: 51
 DB: 15 Gaps: 9

US-10-028-384-12 (1-705) x US-10-028-384-7 (1-2417)

| | | | |
|----|------|--|------|
| QY | 16 | ThrLeuLeuLysLeuLeuLeuLeuSerMetAlaValLeuSerPheSerThrArgLeu | 35 |
| DB | 119 | AGCCTAATCACTTCGCGCATCTCTGTAATCGCCTGGCTGGCGCGGATTTCTCTCGCCTC | 178 |
| QY | 36 | PheAlaValLeuArgPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArg | 55 |
| DB | 179 | TTGCGCGTCATCCGTTTCGATCGATATCCATGAGTTGATCCGTTCAACTACCGG | 238 |
| QY | 56 | ThrThrArgPheLeuAlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArg | 75 |
| DB | 239 | GCCACCGCCTACATGGTGCAGATGGTTGGTACAACTTCCTCAACTGGTTCGACGAGCGC | 298 |
| QY | 76 | AlaTyrPyrProLeuGlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThr | 95 |
| DB | 299 | GCATGGTATCCGCTCGGCGAGATTTGGCGCGTACCCTCTATCCCGCCCTGATGATTAAG | 358 |
| QY | 96 | SerAlaAlaIleTyrHisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal | 115 |
| DB | 359 | TCCGCGCGAATCCATTCGCTGCTGCACGTAACAATACCGGTCATATTCGTGACATC | 418 |
| QY | 116 | CysValPheLeuAlaProLeuPheSerPheThrSerIleValThrTyrLeuLeuThr | 135 |
| DB | 419 | TGGGTGTTCTGGCGCGCATCTTCAGTGGCTGACCTCCATCTCCACTACCTGCTGACC | 478 |
| QY | 136 | LysGluLeuLysAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValPro | 155 |
| DB | 479 | AAGAGGTGTGTCGCGCGCGCGGCTCTTCGCGCCAGCTTCATCGCCATCGTGGCT | 538 |
| QY | 156 | GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys | 175 |
| DB | 539 | GGCTACATCATAGTGGTGGTGGTATCGTACGATAACAGGGCATTTGCCATATTCGCG | 598 |
| QY | 176 | MetLeuLeuThrTyrTyrMetTyrIleLeuAlaValLysThrGlySerIleCysTyrAla | 195 |
| DB | 599 | CTGCAGTTCACTTCTCTGTTGGTGGTCTAGTGAAGCTGATCGGTTCTGGTGG | 658 |
| QY | 196 | AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeu | 215 |
| DB | 659 | GCCGACGCGGTTTGTCTACTTCTACATGGTTCGCGCTTCGCGCTGGCTGGCTGATC | 718 |
| QY | 216 | IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg | 235 |
| DB | 719 | ATCAACCTGATACCCCTGCAGCTTCGTACTGCTCATTTATGGCGAGGTACTCGCGCGT | 778 |
| QY | 236 | IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer | 255 |
| DB | 779 | CTGTGACCACTACACACCTTCTACATCTCGGACTGCTGTCTCTCCATGCGATCCCC | 838 |
| QY | 256 | PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly | 275 |
| DB | 839 | TTGTGGATTTCAACCGATACCCACGACGATGAAACACATGGCTGCGGTGGAGTGTGTG | 898 |
| QY | 276 | LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPhe | 295 |
| DB | 899 | CTCCTATGCGCGTGGCCACTTTCGCGCATTTGCAGTCCGTGCTGTGCGCGCAACGATTC | 958 |
| QY | 296 | GluValLeuPheArgSerValIleSerLeuValGlyPheValLeuThrValGly | 314 |
| DB | 959 | CGAAGCTGTTC-----ATCGTCGCGGATTTGCTGGTGGCGCTTGGCGTC | 1003 |
| QY | 315 | -----AlaLeuLeuMetLeuThr-----GlyLysIleSerProThrPheGlyArgPhe | 330 |
| DB | 1004 | TTTGTGGCGCTGTGGTGTCCACCATGCTGGCGGCTTGTGGCGGCTGTGGAGTGGAGCTTC | 1063 |

| | | | |
|----|------|---|------|
| QY | 331 | TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSer | 350 |
| DB | 1064 | TACTCGCTGGGATACTGGCTAGCCAAAGATCCACATTTCCCATTCATTCGCTGTCG | 1123 |
| QY | 351 | GluHisGlnProThrThrTyrSerTyrTyrPheAspLeuGlnLeuValPheMet | 370 |
| DB | 1124 | GAGCATCAGCCACCACTTGGTTCGTTCTTCTTGATCTGCACATCTGCTGTGCGCC | 1183 |
| QY | 371 | PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIle | 390 |
| DB | 1184 | TTCCAGTGGAGTGGTGTACTGATCAAGCAGATCAACGACGAGGCGGTTTTCGTGTG | 1243 |
| QY | 391 | MetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuVal | 410 |
| DB | 1244 | CTGTAGCCATCAGTGGGTTTACTTCGCTGGTGTGATGGTGGTGTGATGTGTGACCTC | 1303 |
| QY | 411 | AlaProValMetSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMet | 430 |
| DB | 1304 | ACGCGGTGTGTGTCATGCTGGCGGAGTGGCTTTTCGGGACTGTTGGATGTGTCTGTG | 1363 |
| QY | 431 | LysAsnLeuAspIleSerArgPro----- | 438 |
| DB | 1364 | CAAGAGGATTCTCTAAGCGAATGGGCACAGCCATAGCGCAGCCACCGAAGTGGATGA | 1423 |
| QY | 439 | -----AspLysLysSerLys----- | 443 |
| DB | 1424 | GCTGAGGATTCCATTGAGAGAGAGAGCTGTACGACAAAGGCTGGCAAGCTGAAGCATCGT | 1483 |
| QY | 444 | -----LysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGly | 459 |
| DB | 1484 | ACTAAGCATGATGCCGACGAGATACT-----GGCGTCAGCTCCAACCTGAAGAGTATT | 1537 |
| QY | 460 | MetIleLeuValMetAlaPhePheIleThrTyrThrPheHisSerThrTyrValThr | 479 |
| DB | 1538 | GTTATTTGGCGGTTCTAATGCTGTGATGATGTTGCTGCTCCACTGACCGTGGGTGACC | 1597 |
| QY | 480 | SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGly--AspGlySer | 498 |
| DB | 1598 | AGCAATGCTACTTCCAGTCCCTCCATTCTTGGCTTCCACACAGCTCAAGATGGATCC | 1657 |
| QY | 499 | ArgIleIlePheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProGlu | 518 |
| DB | 1658 | CCCAACATTTTAGACGATTCAGAGAGCTTACTACTGGCTTCCGCAACACACTGCCGAT | 1717 |
| QY | 519 | AspAlaLysValMetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg | 538 |
| DB | 1718 | GATGCTCCGTTATGCTTGTGGATACCGATACCGATAGCGGGAATGGCAACAGA | 1777 |
| QY | 539 | ThrIleLeuValAspAsnAsnThrTyrPheAsnAsnThrHisIleSerArgValGlyGlnAla | 558 |
| DB | 1778 | ACGACGCTAGTGGATAATAATACGTGGAACTAGTACATAGTCACATAGCGCTGGTGGCAAGCA | 1837 |
| QY | 559 | MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal | 578 |
| DB | 1838 | ATGCTTCAACCGAGGAGAGTCTCTACAAATTATGATCATCTTGTGAGTGGATACGTT | 1897 |
| QY | 579 | LeuValIlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTyr | 598 |
| DB | 1898 | TTGTGTATCTTGGCGGTGATCGGCTATTCTGCGATGATATCAACAAGTTCCTCTGTGG | 1957 |
| QY | 599 | MetValArgIleGlyGlySerThrAspThrGlyLysHis-----IleLysGluAsn | 615 |
| DB | 1958 | ATGTCCCAATTGCT-----GAGGAGAGCATCCCAAGGACATTAAGGAAGAC | 2005 |
| QY | 616 | AspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeu | 635 |
| DB | 2006 | GATTACTTACCGACCGCGGTGAATTCAGGGTAGATCCGAGAGGTGCTCCGGCCCTGCTC | 2065 |
| QY | 636 | AsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLys | 655 |
| DB | 2066 | AACGCGCTTATGTACAAATTAAGCTACTACAGATTCGGGAAATTAAGATTTGGACTACAGA | 2125 |
| QY | 656 | ArgProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeu | 675 |

us-10-028-384-12.rnpb

Thu Dec 16 16:25:00 2004

| | | | |
|----|------|---|------|
| Db | 2126 | GGTCCATCTGGATATGATCGCACACGTAAACCCGTCATTTGGGAATAAGGACTTCGATCTG | 2185 |
| Qy | 676 | AspValLeuGluGluGlyThrThrGluHisTrpLeuValArgIleTyrLysValIys | 695 |
| Db | 2186 | ACCTACCTGGAGGAGGCCTACACCACAGACACTGGCTTGTTCGCACTATAGGTGAAG | 2245 |

Search completed: December 15, 2004, 21:20:06
Job time : 981.605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 18:52:55 ; Search time 5457.63 Seconds
(without alignments).
4707.175 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 3697

Sequence: 1 MTKGFLRLSYEKQDTLLK.....HMLVRIYKVKLDNRLGSLRT 705

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgr2.1/USPTO.spool/US10028384/runat_14122004_131519_8030/app_query.fasta_1.3740
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -IOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=est -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=est -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@CGN_1_13960@runat_14122004_131519_8030 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST :
1: gb_est1 :
2: gb_est2 :
3: gb_est3 :
4: gb_est4 :
5: gb_est5 :
6: gb_est6 :
7: gb_est7 :
8: gb_ges1 :
9: gb_ges2 :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|------------|--------------------|
| 1 | 3622 | 98.0 | 3802 | 3 AK030363 | AK030363 Mus muscu |
| 2 | 3620 | 97.9 | 2730 | 3 AK081547 | AK081547 Mus muscu |
| 3 | 3605 | 97.5 | 3643 | 3 AK077877 | AK077877 Mus muscu |
| 4 | 3600.5 | 97.4 | 2656 | 3 BC028897 | BC028897 Mus muscu |
| 5 | 3495 | 94.5 | 2118 | 9 AY418284 | AY418284 Homo sapi |
| 6 | 3428.5 | 92.7 | 2103 | 9 AY418286 | AY418286 Mus muscu |
| 7 | 3419.5 | 92.5 | 2727 | 3 AK084770 | AK084770 Mus muscu |
| 8 | 2674 | 72.3 | 2079 | 9 AY418285 | AY418285 Pan trogl |
| 9 | 2239.5 | 60.6 | 4513 | 3 BC044321 | BC044321 Xenopus 1 |

| | | | | | |
|----|--------|------|------|------------|----------------------|
| 10 | 2237 | 60.5 | 2709 | 3 AK018758 | AK018758 Mus muscu |
| 11 | 2201 | 59.5 | 2669 | 3 AK012153 | AK012153 Mus muscu |
| 12 | 2199 | 59.5 | 2419 | 3 CNS0ADYZ | EX816490 Arabidops |
| 13 | 1894 | 51.2 | 1345 | 3 CR606556 | CR606556 full-length |
| 14 | 1740 | 47.1 | 1145 | 5 BX446691 | BX446691 BX446691 |
| 15 | 1706 | 46.1 | 1124 | 7 CK030365 | CK030365 AGENCOURT |
| 16 | 1671.5 | 45.2 | 2334 | 3 AY310157 | AY310157 Rattus no |
| 17 | 1643 | 44.4 | 978 | 7 CR413615 | CR413615 CR413615 |
| 18 | 1539 | 41.6 | 1141 | 6 CD505084 | CD505084 CDA71-G02 |
| 19 | 1514 | 41.0 | 960 | 7 CN062150 | CN062150 Ag2_P35_L |
| 20 | 1455 | 39.4 | 1019 | 5 BX404479 | BX404479 BX404479 |
| 21 | 1447 | 39.1 | 919 | 7 CR439276 | CR439276 CR439276 |
| 22 | 1444 | 39.1 | 918 | 5 BX369502 | BX369502 BX369502 |
| 23 | 1433 | 38.8 | 1037 | 4 BM451400 | BM451400 AGENCOURT |
| 24 | 1418 | 38.4 | 849 | 7 CR409802 | CR409802 CR409802 |
| 25 | 1414 | 38.2 | 863 | 5 BX741571 | BX741571 BX741571 |
| 26 | 1404 | 38.0 | 1185 | 6 CD505085 | CD505085 CDA71-G02 |
| 27 | 1392 | 37.7 | 888 | 6 CD243529 | CD243529 AGENCOURT |
| 28 | 1378.5 | 37.3 | 1086 | 5 BU174083 | BU174083 AGENCOURT |
| 29 | 1377 | 37.2 | 831 | 7 CO424065 | CO424065 UT-M-HUO- |
| 30 | 1376.5 | 37.2 | 893 | 5 BX327064 | BX327064 BX327064 |
| 31 | 1356.5 | 36.7 | 867 | 5 BU253920 | BU253920 603414953 |
| 32 | 1354 | 36.6 | 862 | 5 BX713937 | BX713937 BX713937 |
| 33 | 1354 | 36.6 | 863 | 7 CR440471 | CR440471 CR440471 |
| 34 | 1354 | 36.4 | 892 | 7 CK407659 | CK407659 AUF-1LTV |
| 35 | 1344 | 36.2 | 1154 | 4 BM542576 | BM542576 AGENCOURT |
| 36 | 1339 | 36.2 | 1154 | 4 BM542576 | BM542576 AGENCOURT |
| 37 | 1335.5 | 36.1 | 907 | 6 CB589879 | CB589879 AGENCOURT |
| 38 | 1317.5 | 35.6 | 1017 | 5 BU458020 | BU458020 603367101 |
| 39 | 1314.5 | 35.6 | 933 | 5 BU913073 | BU913073 AGENCOURT |
| 40 | 1313 | 35.5 | 860 | 6 CA453997 | CA453997 AGENCOURT |
| 41 | 1310.5 | 35.4 | 816 | 5 BU200678 | BU200678 603101974 |
| 42 | 1310.5 | 35.4 | 915 | 5 BX431319 | BX431319 BX431319 |
| 43 | 1310.5 | 35.4 | 935 | 6 CB208848 | CB208848 AGENCOURT |
| 44 | 1309.5 | 35.4 | 1065 | 4 BM553035 | BM553035 AGENCOURT |
| 45 | 1307 | 35.4 | 903 | 6 CA487501 | CA487501 AGENCOURT |

ALIGNMENTS

AK030363 3802 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male pituitary gland cDNA, RIKEN full-length
enriched library, clone:530402C17 product:OLIGOSACCHARYL
TRANSFERASE STT3, SUBUNIT HOMOLOG (B5) (INTEGRAL MEMBRANE PROTEIN 1)
homolog [Mus musculus], full insert sequence.

ACCESSION AK030363.1 GI:26326354
VERSION HTCC; CAP trapper.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kiseunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,


```

/clone="6030401L06"
/sex="male"
/tissue="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
93..2209
misc_feature
  /note="OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT HOMOLOG
  (B5) (INTEGRAL MEMBRANE PROTEIN 1) homolog [Mus musculus]
  (SWISSPROT|P46978, evidence: FASTV, 99.7%ID, 100%length,
  match=2115)
  putative"

```

misc_feature

ORIGIN

Alignment Scores:

```

Pred. No.: 0 Length: 3643
Score: 3605.00 Matches: 693
Percent Similarity: 98.72% Conservatives: 3
Best Local Similarity: 98.30% Mismatches: 9
Query Match: 97.51% Indels: 1
DB: 3 Gaps: 0

```

US-10-028-384-12 (1-705) x AK077877 (1-3643)

```

QY 1 MethThyLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLysLeu 20
DB 93 ATGACTAAGCTGGGATTTTTCGATTTGCTATGAGAGCAGGACACACTTCTTAAGCTT 152
QY 21 LeuileLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 153 CTCATCTCTCGATGGCTGCTGTGTATCTTTTCTACTCGTCTTTTGTGCTGAGTA 212
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
DB 213 TTTGAAGTGTGCATCCATGAGTTTGCATCCGACTTTAATATCGGACTACCCGGTTTCTG 272
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTyrTrpLeu 80
DB 273 GCTGAGGAGGGTTTATTAATTCATTAACCTGTTGATGACCGGGCTTGGTACCCCTTG 332
QY 81 GlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 333 GGCCGAATCATTGGAGGAACAATTTACCACAGGTTTAATGATCACTTCCTGCTGCAATCTAC 392
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 393 CATGACTCCATTTCTTCATATACCATTCAGATTCAGATTCGATGCTGGGTATATTTCTGACC 452
QY 121 ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
DB 453 CCGCTTTTCTCTCTCTTACCACCATCGTTACGTACCACTTACCAGAGCTCAAGGAT 512
QY 141 AlaGlyAlaGlyLeuLeuAlaAlaValIleAlaValProGlyTyrIleSerArg 160
DB 513 GCAGGAGCTGGGCTTCTCTCTGCTGCCATGATGCTGATGCTGCTGGGTATATTTCTCGA 572
QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
DB 573 TCTGTAGCTGGCTCTCTCTATTAATGAAGAAATGTCTATCTTTGATGCTGCTTACTTAC 632
QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
DB 633 TACATGTGATCAAGCAGTGAAGACTGGTTCCATCTATTTGGCTGGCCAAAGTGGCCCTC 692
QY 201 AlaTyrPheTyrMetValSerTrpGlyGlyTyrValPheLeuLeuAsnLeuIlePro 220
DB 693 GCTTATTTCTACATGCTCTCTTCTGAGGAGCTATGTTTCTCTGATCAACTTGATTCCT 752
QY 221 LeuHisValLeuValLeuValLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
DB 753 CTACATGCTCGTGTGTAAATGCTGACAGCCGCTTTCTCAGCGGATCTACGAGCTTAC 812
QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260

```

```

DB 813 TGTACTGTTTACTGCTGGGACCACTTCTTCTATGAGATTTCCTTTGTGTTCCAG 872
QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
DB 873 CCCGTCCTTTTCATCAGAACACATGGCAGCCTTTGGAGTGTGTGCTCTCTCAGATCCAT 932
QY 281 AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPheArg 300
DB 933 GCTTTCGTAGATTACCTCGCAGCAAGTTGAATCCACAGCAATTCGAAGTCTCTTTCCGG 992
QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
DB 993 AGTGTATCTCTCGTGGTTGCTTGTCTCTCTCACTGTGGAGCTCTCTCATGCTAACCA 1052
QY 321 GlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
DB 1053 GGAATAATTTCTCCCTGGACAGGCGTTTCTACTCTCTGCTGGATCCCTCTTATGCTAAG 1112
QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
DB 1113 AATAACATTTCCATTTATTTGCACTCTGTTTCGAGCACCAGCCCAACCTGCTCTTCTAC 1172
QY 361 TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
DB 1173 TATTTTGAATCTACAGCTCTTGTCTTCTATGTTTCCAGTTGGCTCTTATTTACTGCTTAGC 1232
QY 381 AsnLeuSerAspAlaArgGlyPheIleIleMetTyrGlyValThrSerMetTyrPheSer 400
DB 1233 AACCTGTCTGATGCTCGGATTTTATCATCATGATGATGTCAGCAGCACTATTTTCA 1292
QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
DB 1293 GCTGTAATGGTGGCTCTAATGCTGTAATGGCACTGTTATGTCATTTCTTTCTGGCAT 1352
QY 421 GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys 440
DB 1353 GGTGTTTCCAGGCTGCTGTCCACATATATGAATAATCTGGACATAGTCTGCCGACAGCA 1412
QY 441 LysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet 460
DB 1413 AGAGCAGAACACACAGGATTTCTACTTACCCTATTAAGAATGAGTGGCGAGTGGGATG 1472
QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
DB 1473 ATACTGTCATGGCTTCTTCTCATCACCTACAGTTTCTTCTTCTTCTTCTTCTTCTTCT 1532
QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaAAGGlyGlyVAspGlySerArgIle 500
DB 1533 GAAGCTTATTTCTTCTCCCTCCATTTGTTCTGCTGCTCGTGGTGGGATGGCAGTAGGATC 1592
QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
DB 1593 ATTTTGTAGTACTTCGAGAGCGTATTATTGGCTCCGTCACAAATCTCCAGAGGATGCA 1652
QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle 540
DB 1653 AAAGTCATGTCATGGGAGTATTATGGCTACCAAAATCTGCAATGGCAATCGGAA-ATT 1711
QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
DB 1712 TTAGTGGCAATAACACATGAATAATATCCCATATTTCTCGAGTAGGGCAGCAATGGCA 1771
QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
DB 1772 TCACAGAGAGAAAGAACCCATGAATCATGAGGAGCTTGATGTCAGCTATGCTGCTGTC 1831
QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
DB 1832 ATTTTGGAGGCTTACTGGTATTTCTCGATGATATCAACAAGTTTCTTTGGATGGTC 1891
QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro 620
DB 1892 CGGATTTGGAGGAAGCAGAGCAGGAGACACATTAAGAGAGATGACTACTATCTCTCT 1951

```

QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuLeuLeuLeuLeuMetTyr 640
 Db 1952 ACTGGGAATTCCTGGTGTATGTCGAGGGTCTCCGGTGTCTCAACCTTATGTAC 2011

QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 Db 2012 AAAATGTGTACTACCGCTTTGGCAGGTCTACACAGAGCCAGCGTCCACAGGCTTT 2071

QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuLeuValLeuGluGlu 680
 Db 2072 GACCGTGTTCGAATGTCAGATGGTAATAGACTTTGAGTTGATCTCTGGAGAA 2131

QY 681 GlyTyrThrGluHisTyrPheLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2132 CGGTATACCAAGAACACTGCTAGTCAGGATATACAAAGTAAGGACCTGGATATCGA 2191

QY 701 GlyLeuSerArgThr 705

Db 2192 GCCTGTGCAGGACA 2206

RESULT 4
 BC028897 2656 bp mRNA linear HTC 04-AUG-2004
 LOCUS Mus musculus cDNA clone IMAGE:4168013, containing frame-shift errors.

ACCESSION BC028897

VERSION BC028897.1 GI:22137711

KEYWORDS HTC.

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE
 AUTHORS Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 2656)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bat, N.K., Hopkins, R.F., Jordan, B., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Narusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., Worley, K.J., Malek, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Winkler, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Bouffard, G.G., Blakesley, R.W., Young, A.C., Shevchenko, Y., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smallos, D.E., Schnerch, A., Schein, J.B., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

TITLE
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 12477932

REFERENCE 2 (bases 1 to 2656)

AUTHORS Director MGC Project.

TITLE Direct Submission

JOURNAL Submitted (01-MAY-2002) National Institutes of Health, Mammalian
 GENE Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: ccapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LNL)

Sequencing Center by: Baylor College of Medicine Human Genome

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clon distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 24 Row: f Column: 2
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 31560679
 This clone has the following problem: frame shifted.

FEATURES

source

1. 2656
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4168013"
 /tissue_type="Salivary gland, 10 week old female mouse"
 /clone_lib="NCI_CGAP_SG2"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 2656
 Score: 3600.50 Matches: 633
 Percent Similarity: 98.72% Conservative: 3
 Best Local Similarity: 98.30% Mismatches: 9
 Query Match: 97.39% Indels: 1
 DB: 3 Gaps: 0
 US-10-028-384-12 (1-705) x BC028897 (1-2656)
 QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 Db 20 ATGACTAAGCTTGGATTTTTCGATGTTCTCTATGAGAGCAGGACACACTTTCTAAGCTT 79
 QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 Db 80 CTCATCTCTGTCGATGCTGCTGTTATCTTTTCTACTCGTCTTTTGTCTGTCTGAGA 139
 QY 41 PheGluSerValLeuHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 140 TTTGAAAGTCTCATCCATGATGTTGATCGTACTTTTAAATATTCGGACTACCCGGTTCTG 199
 QY 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspAspArgAlaTyrProLeu 80
 Db 200 GCTGAGGAGGGGTTTATAAATTCATTAAGTGTGATGACCGGGCTGGTACCTTTG 259
 QY 81 GlyArgIleleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
 Db 260 GGCGGAATCATTCGAGGAACAATTTACCAGGTTTAAATGATCACTTCTGCTCAATCTAC 319
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 320 CATGTACTCCATTTCTCCATATCATTAATGACATTCGGAATGCTGTGTTCTCTGGCC 379
 QY 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
 Db 380 CCACCTTTCTCTCTCTTTCACCCATCGTTAGCTACCCCTTACCAGAGCTCAAGAT 439
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
 Db 440 GCAGGAGCTGGCTTCTTGTCTGCTGCCATGATTCGTGTAGTTCTCTGGGTATATTCTCGA 499
 QY 161 SerValInlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 Db 500 TCTGTAGCTGCTCCTCATGATATAAAGGAATTCGTATCTTTTGTATGCTGTCTACTTAC 559
 QY 181 TyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu 200
 Db 560 TACATGTGGATCAAGGACGTGAAGCTGGTTCATCTATTGGGCTGCCCAAGTGTGCCCTC 619

| | | | |
|----|------|---|------|
| QY | 201 | AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIlePro | 220 |
| Db | 620 | GCTATTTCACATGGTCTCTCATGGGAGGCATGTCTCCTGATCAACTTGATCCCT | 679 |
| QY | 221 | LeuHisValLeuValLeuMetLeuThrGlyAArgPheSerHisAtqGileTyValAlaTyr | 240 |
| Db | 680 | CTACATGTCCTGGTGTAAAGCTGCACAGCGCGTTTTCTCACGGATCTAGTAGCCCTAC | 739 |
| QY | 241 | CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleserPheValGlyPheGln | 260 |
| Db | 740 | TGTACTGTTTACTCCCTGGGCACCATCTCTTATGACAGATTTCCTTTGGTTGCCAG | 799 |
| QY | 261 | ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis | 280 |
| Db | 800 | CCCGTCCTTCATCAGAACAACATGGCAGCCTTTGGAGTCTTTGGTCTCTGTGCAGATCCAT | 859 |
| QY | 281 | AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGlnValLeuPheArg | 300 |
| Db | 860 | GCTTTCGTGAGATTACCTGGCAGCAGTGAATCCACAGAAATCGAAGTCTCTTTCCGG | 919 |
| QY | 301 | SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr | 320 |
| Db | 920 | AGTGTATTCTCCCTGGTGGCTGTGCTCCTCACTGTGGAGCTCTCCTCATGCTAACCA | 979 |
| QY | 321 | GlyLysIleSerProTrpThrGlyAArgPheTyrSerLeuLeuAspProSerTyrAlaLys | 340 |
| Db | 980 | GGAAAAATTTCTCCCTCGACAGGGCGTTTCTACTCTGCTGGATCCCTCTATTGCTTAAG | 1039 |
| QY | 341 | AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr | 360 |
| Db | 1040 | AATAACATTCACATATTGTCATCTGTTCTGAGCACAGCCCACAACTGCTCTTCCTAC | 1059 |
| QY | 361 | TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer | 380 |
| Db | 1100 | TATTTTGATCTACAGCTCCCTGTCTTCATGTTTCCAGTTGGCTCTATTACTGCTTTAGC | 1159 |
| QY | 381 | AsnLeuSerAspAlaAhrgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer | 400 |
| Db | 1160 | AACCCTGCTGANGCTCGAATTTTATCATCATGTATGFGTGCACGACGATGATCTTTTCA | 1219 |
| QY | 401 | AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle | 420 |
| Db | 1220 | GCTGTAATGGTGGCGCTAAATGCTGGTATTGGCACCTGTTATGTGCAATCTTTCTGGCAT | 1279 |
| QY | 421 | GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys | 440 |
| Db | 1280 | GGTGTTTCCCCAGTGCTGTCCACATAG--AAAATCTGGACATAGTGCGCCACAGACAG | 1337 |
| QY | 441 | LysSerLysGlnInAspSerThrTyrProIleLysIleGluValAlaSerGlyMet | 460 |
| Db | 1338 | AAGAGCAAGAAGCAACAGGATCTACTATCCCTATTAAAGAAATGAGTGGCGAGTGGGATG | 1397 |
| QY | 461 | IleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSer | 480 |
| Db | 1398 | ATACTGGTTCATGGCTTTTTTTCATCACCTACACGTTTCATTCGACTTGGGTGACCCAGT | 1457 |
| QY | 481 | GlualaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle | 500 |
| Db | 1458 | GAAGCCTATTCTTCCCTCCAATGACGTCTGTGCTGFGTGGGATGGCAGTAGGATC | 1517 |
| QY | 501 | IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla | 520 |
| Db | 1518 | ATTTTGTATGACTTCCGAGAAGCGTATTATGGCTCCGTCCAAATPATCTCCAGAGATGCA | 1577 |
| QY | 521 | LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle | 540 |
| Db | 1578 | AAAGTCATGTCATGGTGGGATTTAGCTACCAAATPATCTGCMAATGGCAATTCGCACAA | 1637 |
| QY | 541 | LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla | 560 |
| Db | 1638 | TTAGTGGCAAATACATCGAATAATACCATATTTCTCGTAGTAGGACGACGAATGCCA | 1697 |

| | | | |
|------------|---|---|-----------------|
| QY | 561 | SerThrGluGluLysAlaTyGluIleMetArgGluLeuAspValSerTyValLeuVal | 580 |
| Db | 1698 | TCACAGAGAAAAGCCTATGAATCATGAGGAGCTTGATGTACGCTATGCTGTGTC | 1757 |
| QY | 581 | IlePheGlyGlyLeuThrGlyTySerSerAspAspIleAsnLysPheLeuTirMetVal | 600 |
| Db | 1758 | ATTTTTGGAGCCCTTACTGGGTATCTTCGGATGATATCAACAAGTTCTTTGGATGTC | 1817 |
| QY | 601 | ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyTyThrPro | 620 |
| Db | 1818 | CGAGTTGGAGAAACACAGAGACAGAGACACATTAAAGGAGAATGACTACTATACTCT | 1877 |
| QY | 621 | ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTy | 640 |
| Db | 1878 | ACTGGGGAATTCGGTGTGATCGTAGGGTTCTCCGGTGTCTCAACTGCCTTATGTAC | 1937 |
| QY | 641 | LysMetCysTyTyArgPheGlyGlnValTyThrGluAlaLysArgProGlyPhe | 660 |
| Db | 1938 | AAATGTGTTACTACCCCTTTGGCAGGTCTACACAGAGCCACAGGTCACACAGGCTT | 1997 |
| QY | 661 | AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu | 680 |
| Db | 1998 | GACCGTGTTCGAATGCTGAGATGTGGTAAATAAAGACTTTGAGCTGTGATCTCGAGGAA | 2057 |
| QY | 681 | GlyTyThrThrGluHisTyrLeuValArgIleTyLysValLysAspLeuAspAsnArg | 700 |
| Db | 2058 | CGGTATACCACAGAACACTGGCTAGTCAGGATATACAGGTAAGGACCTGGATAATCGA | 2117 |
| QY | 701 | GlyLeuSerArgThr | 705 |
| Db | 2118 | GGCTGTGTCAGGACA | 2132 |
| RESULT 5 | AY418284 | 2118 bp DNA linear | GSS 17-DEC-2003 |
| LOCUS | AY418284 | Homo sapiens ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence, | |
| DEFINITION | AY418284 | genomic survey sequence. | |
| ACCESSION | AY418284.1 | GI:39774244 | |
| VERSION | GSS | | |
| KEYWORDS | Homo sapiens (human) | | |
| SOURCE | Homo sapiens | | |
| ORGANISM | Homo sapiens | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| AUTHORS | 1. (bases 1 to 2118); Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M. | | |
| TITLE | Infering nonneutral evolution from human-chimp-mouse orthologous gene trios | | |
| JOURNAL | Science 302 (5652), 1960-1963 (2003) | | |
| PUBMED | 14671302 | | |
| REFERENCE | 2. (bases 1 to 2118); Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M. | | |
| AUTHORS | Direct Submission | | |
| TITLE | Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA | | |
| JOURNAL | These sequences were made by sequencing genomic exons and ordering them based on alignment. | | |
| COMMENT | Location/Qualifiers | | |
| FEATURES | 1. 2118 | | |
| source | /organism="Homo sapiens" | | |
| gene | /mol_type="genomic DNA" | | |
| | /db_xref="taxon:9606" | | |
| | <1..2118 | | |
| | /gene="ITM1" | | |
| ORIGIN | /locus_tag="HCM6506" | | |

Alignment Scores:

Pred. No.: 0 Length: 2118
 Score: 3495.00 Matches: 670
 Percent Similarity: 95.18% Conservative: 1
 Best Local Similarity: 95.04% Mismatches: 34
 Query Match: 94.54% Indels: 0
 DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AV418284 (1-2118)

QY 1 MetThrIysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 1 ATGACTAAGTTGGATTTCCTGCGATTCCTATGAGAGGAGGACACACATTTTGAAGCTT 60
 QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 61 CTCAATCTGTCATGCGCTGCTGTTATATNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 DB 121 NNN 180
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaIleTyrProLeu 80
 DB 181 GCTGAGGAGGGTTTATAATTCATAACTGGTTGATGACCGAGGCTGGTACCCTTGG 240
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
 DB 241 GGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACTCTGCTGCAATCTAC 300
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 301 CATGTAACCTATTTTCCACATCACCATCGACATTCGGAATGCTGTGTGTTCTCTGGGC 360
 QY 121 ProLeuPheSerPheThrSerIleValThrTyrLeuLeuThrIysGluLeuLysAsp 140
 DB 361 CCT 420
 QY 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArg 160
 DB 421 GCAGGGGCTGGCT 480
 QY 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
 DB 481 TCTGTGGCTGGCTCTCTATGTAATGAAGGATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 540
 QY 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
 DB 541 TACATGTGATCAAGCGAGTAAGACTGGTTCCATCTGTTGGGAGCTAAGTGTGCTCT 600
 QY 201 AlaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuLeuLeuLeuPro 220
 DB 601 GCTTATTTCTCATGGTCTCTCATGGGAGGTTATGTTCTCTGATCAACTTAATTCCT 660
 QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 661 CTCACGCTCTCTGCTGATGCTCACAGGCGGTTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 721 TGATCTGTTTCTGCTCTGGGCACTATATCTTTCTATGACAGATCTCTTTGTGGGTTTCCAG 780
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 781 CCTGTCTTTCTATCAGACACATGCGAGCCTTTGGGGCTTTGGTCTCTCTCTCTCTCTCT 840
 QY 281 AlaPheValAspTyrLeuArgSerIysLeuAsnProGlnPheGluValLeuPheArg 300
 DB 841 GCCTTTGTGGATTACCTCGCAGCAAGTTGAATCCACAACAAATTTGAAGTCTCTTTCCGG 900
 QY 301 SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
 DB 901 AGCGTCATCTCTCTGTAGGCTTTGTCTCTCTACCCGTGGAGCTCTCTCTCTCTCTCTCT 960

QY 321 GlyIysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 DB 961 NGAAAAATATCTCCCTGGAGCGGGCTTCTACTCGCTGCTGGATCCCTCTTATGCTAAG 1020
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrTrpSerSerTyr 360
 DB 1021 AACAACTCCCATCATTTGCTGTCTGAGCATCAGCCCAACCTGGTCTCTCATAC 1080
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1081 TATTTTACCTTCGAGCTCTCTCGCTTCTATGTTTCCAGTTGGCCTCTATTACTGCTTTAGC 1140
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1141 AACCTGTCTGATGCCCGATTTTATCATCATGTATGGTGTGACCATGACTATTCTTCA 1200
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1201 GCTGTATGTTGGTGGCTCTAATGCTAGTGTGGACCTGTTATGTCATTTCTCTGTCAT 1260
 QY 421 GlyValSerGlnValLeuSerThrTyrMetIysAsnLeuAspIleSerArgProAspLys 440
 DB 1261 GGAGTCTCCAGGTGCTGTCACATACATGAAGAACTGACATAAGTCTGTCAGACAAG 1320
 QY 441 LysSerIysLysGlnGlnAspSerThrTyrProIleIysIleGluValAlaSerGlyMet 460
 DB 1321 AAGAGCAAGAGCAACAGGATTCACCTACCTATTAAAGAAATGAAGTGGCAGTGGGATG 1380
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1381 ATACTGTGCTGCTCTCTCTCTCTCATCCATACCTCTTCTCATCACTGGGTGACCACT 1440
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyValAspGlySerArgIle 500
 DB 1441 GAGGCTTACTCTCTCGCTCAATGTACTATCTGCCGCTGGTGGGATGCGATGAGATC 1500
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla 520
 DB 1501 ATATTGATGACTCCGAGAAGCATATTATTGGCTTCGTCATATACTCCAGAGGATGCG 1560
 QY 521 LysValMetSerTrpTrpAspTyrGlyThrGlnIleThrAlaMetAlaAsnArgThrIle 540
 DB 1561 AAGTCTATGCTCTGCTGGGATTTATGGCTATCAGATTACAGCTATGCAACCCGAACTT 1620
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1621 TTAGTGGACAATAACACATGGAATAATACCATATTCTCGAGTAGGCGAGGCAATGGCG 1680
 QY 561 SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1681 TCCACAGAGGAAAAAGGCTATGAGATCATGAGGAGCTCGATGCTAGCTATGCTGGTGC 1740
 QY 581 IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal 600
 DB 1741 ATTTTGGAGGCTCTCTGCTGATGATATCAACAACTTCTTTGGATGGTC 1800
 QY 601 ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrThrPro 620
 DB 1801 CGGATTGGAGGAGACAGATACAGCAACATATCAAGGAGAAATGACTATTATATCTCCA 1860
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1861 ACTGGGAGTTCCGTGGACCGTGAAGGTCTCCAGTGTCTCACTGCTCACTGCTCATGTAC 1920
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProArgIlePhe 660
 DB 1921 AAGATGTACTATCTCTCTGAGCAGGTTTACAGANNNNNNNAGCGTCTCTCAGGCTTT 1980
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680
 DB 1981 GACCGTGTCCGAATGCTGAGATTGGGAAATAAGACTTTTGAGCTTGTATGCTCTGGAGGAA 2040


```

Db      1908  GGATGGTCCGATTGGAGGAGCACAGACAGACAGACACATTAAGGAGAAATGACTACT 1967
Qy      618   yThrProThrClyGluPheArgValAspArgGlySerProValLeuLeuAsnCysL 638
Db      1968  ATACTCTACTGGGAATTCGGTGTGATCGTAGAGGTTCCTCGGTCTCTCAACTGCC 2027
Qy      638  eMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysAspProp 658
Db      2028  TTATGTACAAAATGTGTACTACCGCTTTGGCGAGGTCTACACAGAGCCAGCGTCCAC 2087
Qy      658  roGlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGluLeuAspVal 678
Db      2088  CAGCGTTTGACCGTGTTCGAATGCTGAGATTGGTAAGACTTTGACGCTTGATGCTC 2147
Qy      678  euGluGluGlyTyrThrThrGluHisTrpLeuValArgileTyrLysValLysAspLeuA 698
Db      2148  TGGAGGAAGCGGTATACACAGAACACTGGCTAGTCAGGATATACAGGTAAGGACCTGG 2207
Qy      698  spAsnArgGlyLeuSerArgThr 705
Db      2208  ATATCGAGGGTTGTCAAGGACA 2230

RESULT 8
AY418285
LOCUS
DEFINITION
Pan troglodytes ITM1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
AY418285
VERSION
AY418285.1 GI:39774245
KEYWORDS
GSS.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 2079)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302
2 (bases 1 to 2079)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
JOURNAL
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
These sequences were made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..2079
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>2079
/gene="ITM1"
/locus_tag="HCW6506"

gene

ORIGIN
Alignment Scores:
Pred. No.: 1,77e-284 Length: 2079
Score: 2674.00 Matches: 534
Percent Similarity: 77.20% Conservative: 1
Best Local Similarity: 77.06% Mismatches: 158
Query Match: 72.33% Indels: 0
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AY418285 (1-2079)

```

```

Qy      1   MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db      1   ATGACTAAGTTTGGATTTTGGCATTTGCTCTATGAGAGCAGACACANNCTGAAGCTT 60
Qy      21   LeuileuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db      61   CTCATTCTGTCAATGGGTGCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
Qy      41   PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db      121  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
Qy      61   AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
Db      181  GCTGNNNNNGGGTTTATAAATCCATAACTGGTTTGTATGACCGAGCCCTGGTACCCCTT 240
Qy      81   GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db      241  GGACGAATCATGGAGGANNNTTACCAGGTTTAATGATCACCTCTGTCGCAACTAC 300
Qy      101  HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db      301  CATGTACTCCATTTTTCACATCACCATCGCAATTCGGAATGCTNTGTGTTCCTGGCC 360
Qy      121  ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db      361  CCTCTCTCTCTCTCCACCACCATGTCACGTACACCTTACCAAGAGCTCAAGGAT 420
Qy      141  AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db      421  GCAGGGGCTGGGCTTCTTGCTGTCATGATGCTGATGCTGATGCTGATATCTCCCGA 480
Qy      161  SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db      481  TCTGTGGCTGGCTCTCTATGATTAAGAGGATTCGCATCTTTGCATGCTACTCACCTAC 540
Qy      181  TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaLysCysAlaLeu 200
Db      541  TACATGTGGATCAAGGAGTAAGACTGGTTCCATCTGTGGGACAGTAAAGTGTGCCCT 600
Qy      201  AlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuLeuPro 220
Db      601  GCTTATTTTACATGGTCTGCTCATGGGAGGTATGTTGTTCTGATCACTTAATTCCT 660
Qy      221  LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
Db      661  CTCCAGCTCTCTGCTGCTGATGCTCACAGGCGGTTTCTCTCACCGGATCTATGTGGCTAC 720
Qy      241  CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
Db      721  TGTACTGTTTACTGCTGGGSCACTATCTTCTATGACAGATCTCTTTTGGGTTTCCAG 780
Qy      261  ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
Db      781  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 840
Qy      281  AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArg 300
Db      841  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 900
Qy      301  SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr 320
Db      901  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
Qy      321  GlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
Db      961  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
Qy      341  AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyr 360
Db      1021  AACAAACATCCCCCATCATTTGCTTCTGTGCTGAGCATCAGCCCAACCTGGTCTCATAC 1080

```


| | | |
|------|--|------|
| 361 | TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer | 380 |
| 1081 | TATTTTGACCTNNAGCTCCCTCGCTCTTCATGTTTCAGATTGGCTCTATTACTGCTTNGC | 1140 |
| 381 | AsnLeuSerAspAlaAaGliePheIleIleMetTyrGlyValThrSerMetTyrPheSer | 400 |
| 1141 | AACCTGTCGTGATGCCGGATTTTATCATCATGATGATGGTGTGACACAGCATGTACTTTTCA | 1200 |
| 401 | AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle | 420 |
| 1201 | GCTGTAATGNN | 1260 |
| 421 | GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys | 440 |
| 1261 | GGAGTCTCCCCAGGTNNGTCCNNATACATGAGAANNTCNACATAANNNNNNCANNNAAG | 1320 |
| 441 | LysSerIlysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet | 460 |
| 1321 | AAGAGCAAGAGNNNNNAGGATTCGCCNCCCCTATTAAAGAATGAAGTGGCAAGTGGGATG | 1380 |
| 461 | IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer | 480 |
| 1381 | ATACTGGTCATGCGCTTCTTTCTCATCACTACACTTTCATTCAACTCGGTCGACCACT | 1440 |
| 481 | GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle | 500 |
| 1441 | GAGCGCTACTCTTCTCCGTCCTCAITGTACTATCTGCCCGCTGGTGGGATGGCAGTAGATC | 1500 |
| 501 | IlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAla | 520 |
| 1501 | ATATTGATGACTCCGAGAGAGCATATTATTGGCTTCGTCACTCATCTCCAGAGGATGCG | 1560 |
| 521 | LysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle | 540 |
| 1561 | AAGGTCATNNNCTGGTGGGATTTATGGCTATCAGAITACAGCTATGGCAAAACCGAACAT | 1620 |
| 541 | LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyAlaMetAla | 560 |
| 1621 | TTAGTGACAAATACACATGGAATATATCCCATATTCTCNNNNNNNNNNNGCAATGGCG | 1680 |
| 561 | SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal | 580 |
| 1681 | TCCACAGAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTGAGCTATGTGCTGGTC | 1740 |
| 581 | IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetVal | 600 |
| 1741 | ATTTTTGGAGCCCTCACTGGGTATTCCTCTGATGATATCAACAGTTNCTNNNGATGGTC | 1800 |
| 601 | ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro | 620 |
| 1801 | CGGATTCGGAGGAGNACAGANNNANGCAACATATCAAGAGAGATGACTATTATCTCCA | 1860 |
| 621 | ThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeuMetTyr | 640 |
| 1861 | ACTGGGAGATTCCGTGTGNNCMTGAAGTTCTCCAGTGTCTCTCANNNNNNNNNNNN | 1920 |
| 641 | LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe | 660 |
| 1921 | NNAGCGTCTCCAGCGTTT | 1980 |
| 661 | AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu | 680 |
| 1981 | GACCGTCTCGAAATGCTGAGATTGGGAATAAAGACTTTTGAGCTTGATGTCCTGGAGAA | 2040 |
| 681 | GlyTyrThrThrGluHisTrpLeuValArgIleTyrLys | 693 |
| 2041 | GCATATACCAACAGAACTGGCTGGTCAGAGATATACAAG | 2079 |

| | | |
|-----------|---|-------------|
| VERSION | BC044321.1 | GI:27882209 |
| KEYWORDS | HTC. | |
| SOURCE | Xenopus laevis | |
| ORGANISM | Xenopus laevis (African clawed frog) | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Xenopus. | |
| AUTHORS | 1 (bases 1 to 4513) | |
| TITLE | Klein, S. and Strausberg, R. | |
| JOURNAL | Direct Submission | |
| REMARK | Submitted (17-JAN-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA | |
| COMMENT | NIH-MGC Project Contact: XGC help desk Email: cgaubs-r@mail.nih.gov Tissue Procurement: Dr. Igor Dawid cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA sequencing by: Institute for Systems Biology http://www.systemsbiology.org contact: amadan@systemsbiology.org Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting | |
| FEATURES | Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.llnl.gov Series: IRAC Plate: 94 Row: 1 Column: 19 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein This clone has the following problem: frame shifted. | |
| SOURCE | 1 - 4513 Location/Qualifiers /organism="Xenopus laevis" /mol_type="mRNA" /db_xref="taxon:8355" /clone="IMAGE:4684177" /tissue_type="Embryo, stage 31/32, Xenopus" /clone_lib="NICHD_XGC_Emb4" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" | |

[illegible]

Db 703 GCTGGGCTCAATCCATGGATTTAAACATGTTAAACATAACCGTTCAATACAGAGATGA 762
 Qy 116 CysValPheLeuAlaProLeuPheSerPheThrSerIleValThrValLeuLeuThr 135
 Db 763 TGTGTGTTCTTAGCACAGTATTAGTGGCTTACATCAATCCACTTTCCTGCTCACT 822
 Qy 136 LysGluLeuLysAspAlaGlyValGlyLeuLeuAlaAlaMetIleAlaValPro 155
 Db 823 CGAGAACTGTGGACACCGAGCGGAGCTTCTAGCTCCCTGCTTCATGTGCTAGTGCCA 982
 Qy 156 GlyTyrIleSerArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCys 175
 Db 893 GGATACATATCCAGATCAGTAGCTGGATCAATTCACAAATCAAGGCAATGCTATCTTGGCA 942
 Qy 176 MetLeuLeuThrTyrTyrMetTyrIleValAlaValLysThrGlySerIleCysTyrAla 195
 Db 943 CTGCAGTTTACATATATTTGGGTAAATCTCTGTAANAACAGGCTCTGCTCTGGGCA 1002
 Qy 196 AlaLysCysAlaLeuAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeu 215
 Db 1003 ATAGGTGTGTGTCTTATTTTATATGGTTTCCGCTGGGTGGTTATGTGTTTATC 1062
 Qy 216 IleAsnLeuIleProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArg 235
 Db 1063 ATTAATCTTATCCACTACATGTGTGTACTGTACTGATGCGAGATACACCAAGAGA 1122
 Qy 236 IleTyrValAlaTyrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSer 255
 Db 1123 GTCTACATAGCTTACAGCACATTTTACATGTGGTTTAACTATCATATGCAGATTCCC 1182
 Qy 256 PheValGlyPheGlnProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGly 275
 Db 1183 TTGTGGGATTTACAGCAATACAGACAGTGAACATATGCGAGCTGCGAGGGTCTTTGCC 1242
 Qy 276 LeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPhe 295
 Db 1243 CTGCTGCAAGCTATGCTCTCTCAGTACTTTGAGACAACTAAGTAAGCAGGATTTC 1302
 Qy 296 GluValLeuPheArgSerValIleSerLeu---ValGlyPheValLeuLeuThrValGly 314
 Db 1303 CAGACACTGTTTTCCTCGGTGTATCGTTGGACCTGGAGCTGATTTTCTCACTGTC--- 1359
 Qy 315 AlaLeuLeuMetLeuThrGlyLysIleSerProThrThrGlyArgPheTyrSerLeuLeu 334
 Db 1360 ATCTATTTCATACAGGCTACATTCCTCAATGGAGTGGAGATTTTACTCACATTGG 1419
 Qy 335 AspProSerTyrAlaLysAsnIleProIleIleAlaSerValSerGluHisGlnPro 354
 Db 1420 GACACAGGGTATGCGAAATCCATATCCCATTCATCTGCTGTCGAACATCAACCA 1479
 Qy 355 ThrThrTyrSerSerTyrTyrPheAspLeuGlnLeuValPheMetPheProValGly 374
 Db 1480 ACAACATGGGTCTCTTTTTCATCTGTCATATATAGTATGATACATTTCTGCGAGGA 1539
 Qy 375 LeuTyrTyrCysPheSerAsnLeuSerAspAlaGlyIlePheIleIleMetTyrGlyVal 394
 Db 1540 CTTTGGTTCTGCATAAAAATATCAACATGAAAGATTTTGTGCTGTGTATGCAATT 1599
 Qy 395 ThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
 Db 1600 AGTGGCGTTTACTTGTGGGTAATGTTGTTGTTGATGCTGCTGCTGCTGCTGTT 1659
 Qy 415 SerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAsp 434
 Db 1660 TGCATGCTGTCGGCAATGCTCTTCCCAATGTTTTTGAGCATTTATCTGGGTGAT---GAT 1716
 Qy 435 IleSerArgPro----- 438
 Db 1717 ATGAAGCGGGAACCCCGCTGGAAGACAGCAGTACGAGAGCGAATTTCTGGTACA 1776
 Qy 439 -----AspLysSerLysGlnGlnAspSerThrTyrProIleLysIleGlu--- 455
 Db 1777 CTCTATGATAGGCTGGCAATGTGAGAAAGCATGTGTCGAGCAGGAGAAAGCAGAGAA 1836

Qy 456 -----ValAlaSerGlyMetIleLeuValMetAlaPhePheLeuLeuThr 470
 Db 1837 GGCTTGGCCCTAATAATAAGAGTATAGTCACTATCTTGTATGTAATGCTGTTGATG 1896
 Qy 471 TyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSerIleValLeu 490
 Db 1897 TTTCTGTACACTGTAGTGGTAACTAGCAATGCTTATCCAGCCCTAGTGTGTTATG 1956
 Qy 491 SerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyr 510
 Db 1957 CGGTATACATACGATGCAACGAAATATCTTGGATGACTTTCAGAGAAGCATACTAC 2016
 Qy 511 TrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTyrAspTyrGlyTyr 530
 Db 2017 TGGTAAAGGAGATACAGATGAACTGCGCGTGTATGCTCTGCTGGGATTTAGGCTAC 2076
 Qy 531 GlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnLeuThr 550
 Db 2077 CAAATTCAGGAATGGCTAATCGAACCCACTTTAGTAGACAACAATACGTGGCAACATAGC 2136
 Qy 551 HisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMet 570
 Db 2137 CACATAGCTTGGTTGAAAGCAATGTCTTCAATGAAACCCGCTCTTATGAAATATG 2196
 Qy 571 ArgGluLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSer 590
 Db 2197 AAAAGCTTGGATGTAGATTTATGTAATAATATTTGGGGTGTAAATGGATATCTGTT 2256
 Qy 591 AspAspIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGlyLys 610
 Db 2257 CATGCATCAACAATTTCTCTGGATGGTTCGAATAGCA-----GNAAGAGAA 2304
 Qy 611 His-----IleLysGluAsnAspTyrTyrThrProThr---GlyGluPheArgValas 627
 Db 2305 CATCAAAAGATATACGGAAGAGTCACTTCACTC-----ACAAGGAGAGTTTCTGTAGA 2361
 Qy 627 PArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPh 647
 Db 2362 CAAGCTGGATCTCCACCTTGTCAATGCTCTCATGTATAGATGTCATATACCGTT 2421
 Qy 647 eGlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaG 667
 Db 2422 TGGTGAATCAGCTTGATTTCCGAAACACCCCTGGATTTGACCGACACGTAATGCTGA 2481
 Qy 667 uIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyrThrThrGluHisTr 687
 Db 2482 AATTGGGATTAAGATTAATAATCAAGCATTTGGAGAAGCATTTACATCGAGGACTG 2541
 Qy 687 pLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 Db 2542 GCTTGTAGATAACAAAGTTAAAAAGCTAGAAAAACAGA 2581

RESULT 10

AK018758

LOCUS

DEFINITION

MUS MUSCULUS ADULT MALE LIVER CDNA, RIKEN FULL-LENGTH ENRICHED

LIBRARY, CLONE:130006C19 PRODUCT:HYPOTHETICAL OLIGOSACCHARYL

TRANSFERASE (OTASE) STT3 SUBUNIT CONTAINING PROTEIN, FULL INSERT

SEQUENCE.

ACCESSION

AK018758

VERSION

AK018758.2

KEYWORDS

HTC; CAP TRAPPER.

SOURCE

MUS MUSCULUS

ORGANISM

EUKARYOTA; EUTHERIA; RODENTIA; SCIURIDAE; MURINAE; MUS.

REFERENCE

1

CARNICI, P. AND HAYASHIZAKI, Y.

HIGH-EFFICIENCY FULL-LENGTH CDNA CLONING

METH. ENZYMOL. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

| FEATURES | source |
|--------------|---|
| AUTHORS | Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishire, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, Y., Kira, A. and Hayashizaki, Y. |
| TITLE | OKAZAKI, Y., MURAMATSU, M., INOUE, Y., KIRA, A. and HAYASHIZAKI, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer |
| JOURNAL | Genome Res. 10 (11), 1757-1771 (2000) |
| MEDLINE | 20530913 |
| PUBMED | 11076861 |
| REFERENCE | 4 |
| AUTHORS | THE RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. |
| TITLE | Functional annotation of a full-length mouse cDNA collection |
| REFERENCE | Nature 409, 695-690 (2001) |
| AUTHORS | 5 |
| TITLE | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. |
| JOURNAL | Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs |
| REFERENCE | Nature 420, 563-573 (2002) |
| AUTHORS | 6 (bases 1 to 2669) |
| TITLE | Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, K., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Soabe, Y., Suzuki, H., Tagawa, M., Tagawa, A., Takahashi, P., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. |
| JOURNAL | Direct Submission |
| REFERENCE | Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) |
| COMMENT | Please visit our web site (http://genome.gsc.riken.jp/) for further details |
| FEATURES | <p>Host: S01R.</p> <p>Location/Qualifiers</p> <p>1..2669</p> <p>organism="Mus musculus"</p> <p>mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="FANTOM DB:2610524N02"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="2610324N02"</p> <p>/tissue_type="whole body"</p> <p>/clone_lib="RIKEN full-length enriched mouse cDNA library"</p> <p>/dev_stages="10 days embryo"</p> <p>36..2507</p> <p>note="hypothetical Oligosaccharyl transferase (OTase) SRT3 subunit containing protein (InterPro IPR003674, evidence: InterPro)"</p> |
| misc_feature | 36..2507 |

QY 295 eGluValLeuPheArgSerValIleSerLeu---ValGlyPheValLeuLeuThrValGI 314
 DB 1073 CCAGACCCCTTTCTTTTGGGTGCTCTACTAGCTGCAGGGCGTGTCTCTAGTGC-- 1130
 QY 314 yAlaLeuLeuMetLeuThrGlyLysIleSerProIleThrGlyArgPheTyrSerLeuLe 334
 DB 1131 -ATCTATCTGCATACACAGCTTATATTGCACCATGGAGTGGCAGGTTTATTCACTATG 1189
 QY 334 uAspProSerTyrAlaLysAsnIleProIleIleAlaSerValSerGluHisGlnPr 354
 DB 1190 GGATCTGGGTATGCAAAATACACATTCCTCAATATTGTCATCATGTGTCTGAACATCAGCC 1249
 QY 354 oThrThrTrpSerTyrPheAspLeuGlnLeuLeuValPheMetPheProValGI 374
 DB 1250 TACGACATGGGTGCTTCTCTCTTGTATACATATTCTGTATGTACCTTCCACAGCG 1309
 QY 374 yLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyVa 394
 DB 1310 CCTATGGTCTGCATCAAAATATACAGCATCAAGAGATATTGTCTGCTGTATCGCAT 1369
 QY 394 lThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMe 414
 DB 1370 CAGTCTGCTACTTTCGGGAGTGTGTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1429
 QY 414 tSerIleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAs 434
 DB 1430 CTGCATGCTGTGGCCATCGCTCTCTCAATGTTTTGAGCAGCATTTTGGGGGT---GA 1486
 QY 434 pIleSerArg-----ProAspLysLysSerLysGlnAspSerThrTyrPr 451
 DB 1487 CATGAAGAGGAAACCCACCTGTCGAGGACGAGTCATGAGTATGACAAAGAACCC 1546
 QY 451 oIleLysIleGluValAlaSerGlyMetIle----- 461
 DB 1547 AGGAAACTTGTATGACAAAGGAGGTAAAGTGAGGAGCATGTGACAGAGCAAGAAACC 1606
 QY 462 -----LeuValMetAlaPhePheLe 468
 DB 1607 TGAAGAGGGCTTGGCCCAACATCAAAAGCATTTGTCACCATGCTGATGCTCATGCTCT 1666
 QY 468 uIleThrTyrThrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSerIl 488
 DB 1667 GATGATGTTGGGGTCCACTGCGTGGGTGTCACAGCAAGCCCTACTCCAGTCCAGTGT 1726
 QY 488 eValLeuSerAlaArgGlyLysArgIlePheAspAspPheArgGluAl 508
 DB 1727 GGTCTTGCCTCTCAATCATGATGTTGACAGGATATATTAGATTTTAGAGAGC 1786
 QY 508 aTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTy 528
 DB 1787 GTACTTTTGGCTTGAGACAAACACCGATGAACACGCGGGTCTATGTCGGTGGGACTA 1846
 QY 528 rGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAs 548
 DB 1847 CGGCTATCAGATGCTGGCATGGCCACAGGACCCACTCTGTGTGATACACACCTGGAA 1906
 QY 548 nAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGI 568
 DB 1907 CAACAGCCACATCGCACTGGTGGGAAAGTATGCTTCTCAATGAAACGCGCCCTATAA 1966
 QY 568 uIleMetArgGluLeuAspValSerTyrValLeuValIlePheGlyLeuThrGlyTy 588
 DB 1967 ATCATAGGTCCTTGTATGTCGATATGTTGTTATTTTCGAGAGATGATGGCTA 2026
 QY 588 rSerSerAspAlaIleAsnLysPheLeuTrpMetValArgIleGlySerThrAspTh 608
 DB 2027 TTTCCGGGGACGATATCAACAGTCTCTGATGTCAGGATAGCT-----GA 2074
 QY 608 rGlyLysHis-----IleLysGluAsnAspTyrTyrThrProThrGlyGluPheAr 625
 DB 2075 AGGGAGCATCCCAAGACATCCGGAGAGGTGACTATTTCACCGAGGAGGAGTTCCG 2134
 QY 625 gValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTy 645

DB 2135 AGTAGAAGAGCTGGGTCTCTCTACTCTGTGTAAACTGCCTTATGTATAAATGTCATACTA 2194
 QY 645 rArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPheAspArgValArgAs 665
 DB 2195 CAGATTTGGAGAAATGACGCTAGATTTTCGCACTCCCGCAGGCTTTGACCGAACAGCTAA 2254
 QY 665 nAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGluGlyTyrThrTrGI 685
 DB 2255 TCTGATGATTGGAATTAAGACATTAATTCAGACATTTGGAGGAGCTTTTACATCAGA 2314
 QY 685 uHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg 700
 DB 2315 GCACCTGGCTGTGAGGATATATAAAGTGAAGCAACCTGACACAGG 2360

RESULT 12
 CDSADYZ CDSADYZ 2419 bp mRNA linear HNC 06-FEB-2004
 LOCUS Arabidopsis thaliana Full-length cDNA Complete sequence from clone
 DEFINITION GSLTPGH512C09 of Hormone Treated Callus of strain col-0 of
 Arabidopsis thaliana (thale cress).
 ACCESSION BX816490.1 GI:42474670
 VERSION HNC: GSLT cDNA
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 2419)
 AUTHORS Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C.,
 Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V.,
 Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.
 TITLE Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
 A Combined Approach to Evaluate and Improve Arabidopsis Genome
 Annotation
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2419)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (19-NOV-2003) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr
 The sequences are based on single pass reads.
 Life Technologies (a division of Invitrogen) members carried out
 full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castelli
 V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C.,
 Schachter V., Weissenbach J., Salanoubat M.
 URV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis
 genome released by MIPS (Munich Information center for Protein
 Sequences). 5 prime and 3 prime are assembled with Phrap.
 http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
 length
 http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
 FEATURES
 source Location/Qualifiers
 1..2419
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLTPGH512C09"
 /tissue_type="Hormone Treated Callus"
 /plasmid="PCMVSPORT_6"
 1..2419
 /gene="Atig34130"

gene
 ORIGIN
 Alignment Scores:
 Pred. No.: 9,5e-232 Length: 2419
 Score: 2199.00 Matches:
 Percent Similarity: 73.68% Conservative: 111

| | | | | | |
|--|------|---|-----------------|--|--|
| Best Local Similarity: 58.22% | | | Mismatches: 157 | | |
| Query Match: 59.48% | | | Indels: 32 | | |
| DB: 3 | | | Gaps: 9 | | |
| US-10-028-384-12 (1-705) x CNS0ADYZ (1-2419) | | | | | |
| QY | 4 | PheGlyPheLeuArgLeuSerTy:GluLysGlnAspThrLeuLeuLysLeuLeuLeu 23 | | | |
| DB | 189 | TTCAATCCCTAAAATTGAAGACGAACAAACAGAGCTTACTCCGATCTCGATCCIT 248 | | | |
| QY | 24 | SerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSer 43 | | | |
| DB | 249 | GGTCTCGTCTACATCTTAGCTTTCATAGCTCGTCTCTCTAGTGTCTCCGCTACGAATCA 308 | | | |
| QY | 44 | ValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeuAlaGluGlu 63 | | | |
| DB | 309 | ATGATCCAGCAATTCGATCCGATTTCACTACCGTAGACTCTTTCTTCACAGAGAAA 368 | | | |
| QY | 64 | GlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeuGlyArgIle 83 | | | |
| DB | 369 | GGCTTCTACAGTTTGGAACTGGTTCGATTCGAGAGTTGGTATCTCTAGTTCGAATC 428 | | | |
| QY | 84 | IleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeu 103 | | | |
| DB | 429 | ATCGCGGTAATCTTATCCCGCTCTTATGGTCACAGCTGGTAACTACTCTGACCAITG 488 | | | |
| QY | 104 | HisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPhe 123 | | | |
| DB | 489 | CGATTCTCCGATCTTCGTTCCATATCCGTCGAGGTTTGTGTAAACCGCACCGTTTTC 548 | | | |
| QY | 124 | SerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAspAlaGlyAla 143 | | | |
| DB | 549 | CGGTCAATACGACTCTCGTTCGCTATTTCTTGTAAAGAGCTTGGGATACAGGAGCT 608 | | | |
| QY | 144 | GlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSerArgSerValAla 163 | | | |
| DB | 609 | GGTCTGTGCTGCTGTTTATCGCTATTTGCTCGGATACATTTCTGTTCTGTAGCT 668 | | | |
| QY | 164 | GlySerTyAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrp 183 | | | |
| DB | 669 | GGTTCGTATGATAACGAGCTGTGCGATTTTCGCTCTGTTCTGACGTTTATCTCTT 728 | | | |
| QY | 184 | IleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeuAlaTyrPhe 203 | | | |
| DB | 729 | GTAAAGCGGTGAACCTGTTTCAATAGCTTGGCTCTTTCGCTCTGTTCTGCTTACTTT 788 | | | |
| QY | 204 | TyrMetValSerSerTrpGlyTyrValPheLeuLeuAsnLeuProLeuHisVal 223 | | | |
| DB | 789 | TATATGTTTTCAGCTTGGGAGGATATGTGTTTATATCAACTTGGTCTCTTTACGCTG 848 | | | |
| QY | 224 | LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243 | | | |
| DB | 849 | TTGGTGTGTGATACCGGAGGATATCGATGAGCTTTATATGCTTTATTAACGTATG 908 | | | |
| QY | 244 | TyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGlnProValLeu 263 | | | |
| DB | 909 | TATATCTGGGATGTGCTGGGATGCAATTCGCTTGTGTTTC---ACATGTTCA 965 | | | |
| QY | 264 | SerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHisAlaPheVal 283 | | | |
| DB | 966 | TCGTGGTGAACATATGGGTGCGATCGGTGCTCTTATTAATGACGAGTGTTCCTCTG 1025 | | | |
| QY | 284 | AspTyrLeuArgSerLysLeuAsnProGlnGln---PheGluValLeuPheArgSerVal 302 | | | |
| DB | 1026 | GACTGGGTAAATACAGCTCAATGACCAAGTTGTTCCAAACGGTTTGGAGATAACT 1085 | | | |
| QY | 303 | IleSer-----LeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeu 319 | | | |
| DB | 1086 | GTGACATCGGCTATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1136 | | | |
| QY | 320 | ThrGlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAla 339 | | | |
| DB | 1137 | TCGTGGCTATATCTCCATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1196 | | | |
| QY | 340 | LysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSer 359 | | | |
| DB | 1197 | AAGGATCATATCTTATTTATGTCATCTGTTTCTGAGCATCAGCCTACTGCTGCTGCTCA 1256 | | | |
| QY | 360 | TyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPhe 379 | | | |
| DB | 1257 | TTTCACTTTTGAATACCATATTTCTCTTTCTTTCTCTCGGGTCTTTACTTCTGTTTC 1316 | | | |
| QY | 380 | SerAsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPhe 399 | | | |
| DB | 1317 | AGCCGTTTGACGAGTACGATATTTATGCTATGATGCTCTCACTAGCTTGTACTTT 1376 | | | |
| QY | 400 | SerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGly 419 | | | |
| DB | 1377 | GCTGTGTATGTTTGGCTTATTTCTGCTGCTCTCTCAGCAGTTTGTCTTATCAGTGCC 1436 | | | |
| QY | 420 | IleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeu---AspIleSerArgPro 438 | | | |
| DB | 1437 | ATAGCTGTCTCTGCCACTATC-----AAGATTTAACTCTCTCTGTAAGGACA 1484 | | | |
| QY | 439 | AspLysLys-----Ser 442 | | | |
| DB | 1485 | AAACAAAGGTTTCTCAGACTGTTTCCAGAAAGAGCTGGTAGTCAAAAGCTTCTTCA 1544 | | | |
| QY | 443 | LysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMetIleLeu 462 | | | |
| DB | 1545 | AAGTTTACACTTGATCAGCTCAGCTTCCAGAGAATGGTGCCATTGCTCTTCTTGT 1604 | | | |
| QY | 463 | ValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSerGluAla 482 | | | |
| DB | 1605 | GGTGTATTTATTTGCTCAGTAGATATGCTATCTACTGCACATGGGTGACAGAGGCA 1664 | | | |
| QY | 483 | TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502 | | | |
| DB | 1665 | TATTCATCTCCCTCAATGCTTGTAGTCGACAGAGAGCCACCGGAGACAGATCATCTT 1724 | | | |
| QY | 503 | AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal 522 | | | |
| DB | 1725 | GATGATTACCGGAGGACATCTACTTGGCTTAGGCAAAACACTGCCACTGATCTAAGAT 1784 | | | |
| QY | 523 | MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542 | | | |
| DB | 1785 | ATGCTGTGTGGACTATGGTACCAATCATCTGCCATGGGAAATAGAACTGTCTATGTC 1844 | | | |
| QY | 543 | AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562 | | | |
| DB | 1845 | GATACCAATACCTGGAACACACTCATATTGCTACCGTTGGACGCCCATGCTCTTAT 1904 | | | |
| QY | 563 | GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582 | | | |
| DB | 1905 | GAAGATGATCGGTATGACATCATGAGTGCATGTGATGAACTATGTATGTGCTGCTT 1964 | | | |
| QY | 583 | GlyGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrpMetValArgIle 602 | | | |
| DB | 1965 | GGTGTGTACTGTTTATCTTCAGATGATATCAACAGTTCTTGTGGATGTTGAGGATT 2024 | | | |
| QY | 603 | GlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGly 622 | | | |
| DB | 2025 | GGAGGTGGA-----GTATTTCCAGTGTATCAAAAGAACCTGATTACCTGTG---AATGTT 2075 | | | |
| QY | 623 | GluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMet 642 | | | |
| DB | 2076 | GAATTCGCTGTAGCAAAAGGTGATCACCGAAGATGCTGNACTGTCTCATGTACAGTTA 2135 | | | |
| QY | 643 | CysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArg 662 | | | |
| DB | 2136 | TGCTATTACAGTTTCGCTGAGCTGACCACAGAATACCGCAAGACCCCGGCTATGATAGA 2195 | | | |
| QY | 663 | ValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyr 682 | | | |
| DB | 2196 | GCAAGAGAGGTGGAGATAGGGAACAAAGATATCAAACTTGAACACTTAGAAGAGCTTAC 2255 | | | |

COMMENT

On May 22, 2003 this sequence version replaced gi:31027756.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.

This sequence belongs to sequence cluster 6310.f

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna/s=CL0B0092D10RPL&c=6310.f.

FEATURES

Location/Qualifiers

1..1145

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CL0B0092D10"

/tissue_type="NEUROBLASTOMA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:

Pred. No.: 2,15e-181 Length: 1145
Score: 1740.00 Matches: 343
Percent Similarity: 95.56% Conservative: 1
Best Local Similarity: 95.28% Mismatches: 14
Query Match: 47.07% Indels: 4
DB: 5 Gaps: 2

US-10-028-384-12 (1-705) x BX446691 (1-1145)

QY 343 IleProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrTyrPhe 362
DB 1 ATCCCCATATGCTTGTGTGTCGAGCATGACCCCAACCTGCTCTCATCATATTTT 60
QY 363 AspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSerAsnLeu 382
DB 61 GACCTCAGCTCTGCTTTCATGTTTCCAGTTGGCTCTATTACTGTTTAGCAACCTG 120
QY 383 SerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSerAlaVal 402
DB 121 TCTGATGCCCGGATTTTATCATCATGATGTTGTCGACCATGATCTTTTCAGCTGTA 180
QY 403 MetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIleGlyVal 422
DB 181 ATGTCGGTCTATGCTAGTGTGGACCTGTTATGTGCACTTCTCTGGCATGGAGTY 240
QY 423 SerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProAspIlySer 442
DB 241 TCCAGGTGCTGTCACATACATAGAAATCTGGACATAGTCGTCACAGCAAGAGAGC 300
QY 443 LysLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMetIleLeu 462
DB 301 AAGAGCAACAGGATTCACCTACCTATTAAGATGAAGTGGCAAGTGGATGATACCTG 360
QY 463 ValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSerGluAla 482
DB 361 GTCATGGCTTCTTCTATCATCACTACCTTTCATCAACCTGGGTGACCAAGTAGGCC 420
QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
DB 421 TACTCTCTCCGTCCTATGTTACTATCTGCCGTGGTGGGATGGCAGTAGGATCATATTT 480
QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
DB 481 GATGACTTCGAGAGCATATTTATGGCTTCGTTCATAATACTCCAGAGGATCGAAGGTC 540

QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleIleVal 542
DB 541 ATGTCTGTGGGATTATGGCTATCAGATTACAGTATGGCAACCCGAAACAATTTTAGTG 600
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
DB 601 GACATAACACATGATGAATAAACCCATATTTCTCGATAGGCGAGCAATGCGCTCCACA 660
QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
DB 661 GAGGAAAAAGCCCTATGAGATCATGAGGAGCTCGATGTCACTATGTCTGCTGCTATTTT 720
QY 583 GlyGlyLeuThrGlyTyrSerSerAspIleLeuLysPheLeuTrpMetValArgIle 602
DB 721 GGAGGCGCTCAGTGGGTATTCCTCTCATGATATCAACAAGTTTCTTGGATGGTCCGATTT 780
QY 603 GlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrProThrGly 622
DB 781 GGAGGAGGACACAGATACAGGCAACATATCAAGGAGATGACTATTATATCTCAACTGGG 840
QY 623 GluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLysMet 642
DB 841 GAGTTCCGTTGGACCGTGAAGTTTCTCCAGTGTCTCAACTGTCTCATGTACAGATG 900
QY 643 CysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPheAspArg 662
DB 901 TGTACTATCGCTTTGGACAGGTTTACACAGAGCCAGCGTCTCCAGGCTTTGACCGT 960
QY 663 ValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlyTyr 582
DB 961 GTCCGAATTCCTGAGATTGGGAATAAGACTT-GAGCTTGATGTCTCTCGAGGAR---CAT 1016
QY 683 ThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArgGlyLeu 702
DB 1017 ATACACACATTTGGCKKACGRTATT---ANAGTAAGRCCTGGATAT-CGAGGCCGTG 1072

RESULT 15

CK030365

LOCUS

DEFINITION

5', mRNA sequence.

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 1124)

NIH-MGC http://mgc.ncbi.nlm.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs@remail.nih.gov

Tissue Procurement: Len Zon, Harvard

cDNA Library Preparation: Open Biosystems

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM14825 row: m column: 06

High quality sequence stop: 853.

Location/Qualifiers

1..1124

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

FEATURES

Location/Qualifiers

1..1124

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

source

/clone="IMAGE:7052144"
 /tissue type="whole body"
 /lab host="DH10B"
 /clone lib="NIH ZGC 7"
 /note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;
 Bulk tissue was collected from a whole adult individual
 from the Tuebingen strain. 1st strand cDNA was primed with
 a Not I - oligo(dT) primer, double-stranded cDNA was
 cloned into the Not I and EcoRV sites of pExpress-1.
 Library was size-selected for >1 kb fragments and
 normalized. A non-normalized version of this library is
 also available (NIH ZGC 10). Library was constructed by
 Open Biosystems (Huntsville, AL)."

ORIGIN

Alignment Scores:

Pred. No.: 1.22e-177 Length: 1124
 Score: 1706.00 Matches: 337
 Percent Similarity: 94.84% Conservative: 12
 Best Local Similarity: 91.58% Mismatches: 17
 Query Match: 46.15% Indels: 4
 DB: 7 Gaps: 0

US-10-028-384-12 (1-705) x CK030365 (1-1124)

QY 340 LysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSer 359
 Db 1 AAAAACAACATTCCTCATCGCTCTGTCTGTGACATCAGCCACCATGGTCTCTCC 60
 QY 360 TyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPhe 379
 Db 61 TACTACTTTTACCTCCAGCTTCTGGTCTTCATGTTCCAGTTGGCTTTTACTACTGCTTC 120
 QY 380 SerAsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPhe 399
 Db 121 AACAACTCTCTGATGCCAGCATTTTCATCATGTATGGCGTCCACGACATGTACTTC 180
 QY 400 SerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGly 419
 Db 181 TCAGCTGTGATGTGGTCTCTCATGTTGGTCTCGCTCCAGTAATGTGATCTCTGTCTGGC 240
 QY 420 IleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAsp 439
 Db 241 ATTGGAGTCTCTCAGGTCTCTCACCACTTCATGAGAACCTGGATGTAGTCCAGCTGAC 300
 QY 440 LysLysSerLysLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGly 459
 Db 301 AAGAAAACCAAGAGCAGCAGGACTCCACTACCCCAATCAAGAACGAAGTGCACAGTGGG 360
 QY 460 MetIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThr 479
 Db 361 ATGATTCTAGTGTGGTCTCTTCTCTCATCATACATACATCTCCACTCCAGCTGGGTGACC 420
 QY 480 SerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyValAspGlySerArg 499
 Db 421 AGCGAGGGGTATTCTGTCCTCCCTCCATCGTCTCTCTCGCGTGGCGACGCGCGC 480
 QY 500 IleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAsp 519
 Db 481 ATCATCTTTGATGACTTCAGAGAGGCTTATTACTGGCTCAGACACAAACACACAGAGGAT 540
 QY 520 AlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThr 539
 Db 541 GCTAAAGTGTATGCTGGTGGGATTTATGGGTATCAGATAACAGCGATGGCCCAATCGAACG 600
 QY 540 IleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMet 559
 Db 601 ATTCTGGTTGACAATAACACTTGAATAACACTCACATCTCCAGAGTGGGTGAGCAATG 660
 QY 560 AlaSerThrGluGlnLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeu 579
 Db 661 GCCTCCACAGAGGAGAACCTATGAGATTATGCGAGAGCTGGATGTGAGCTACGCTCCTG 720

QY 580 ValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMet 599
 Db 721 GTGATCTTTGGCGGATTGACCGGATATTTCATCTGATGATATTAATAGTTCTCTGGGATG 780
 QY 600 ValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThr 619
 Db 781 GTGCGTATTCGGTGGCAGCAGACACAGGAGGACATATAAAGAGCAGCATATTATACACG 840
 QY 620 ProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMet 639
 Db 841 CCACACTGGAGAGTTTCGTGTGACCGCGAGGGCTCTCTCTGCTCAACTGCTCTCATG 900
 QY 640 TyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGly 659
 Db 901 TACAAGATGTGCTACTATCTGCTTCNGCCAGGTCTACACCGAAGCCAG-CGTCTCTCTGT- 958
 QY 660 PheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 679
 Db 959 TACGAAAGAGTTTCGTATGCTGAGATTGGGCAATAGGACTTTTAACTGGACGTGCTGGAA 1018
 QY 680 GluGlyTyrThrThrGluHisTrpLeuValArgIle-TyrLysValLysAspLeuAspA 699
 Db 1019 AAAAGCTTTAAACTACAAACACACTGGCTGGCGCAGATAATACAGGCCAAGATTTTACA 1078
 QY 699 snAtgGlyLeuSerArgThr 705
 Db 1079 ATCGCGGCCCTTCGAAACC 1098

Search completed: December 15, 2004, 11:54:55

Job time : 5547.63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model.

Run on: December 14, 2004, 13:16:57; Search time 7810.5 Seconds
(without alignments)
4268.513 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTKFGFURLSYEKQDTLLKL.....HMLVRIYKYKDLNRLSLRT 705

Scoring table:

| OLIGO | Xgapop 60.0 | Xgapext 60.0 |
|-------------|--------------|--------------|
| Ygapop 60.0 | Ygapext 60.0 | |
| Fgapop 6.0 | Fgapext 7.0 | |
| Delop 6.0 | Delext 7.0 | |

Searched: 4526729 seqs, 23644849745 residues

Word size: 1

Total number of hits satisfying chosen parameters: 9045947

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10028384/runat.14122004.131645.6800/app.query.fasta.1.3740
-DB=genEmbl -QW=faetap -SUFFIX=Oligo.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -STAR=1 -END=-1 -WAITRX=Oligo -TRANS=human40.cdd -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN.1.1.20262 @runat.14122004.131645.6800 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -XGAPOP=60
-XGAPEXT=7 -XGAPOP=60 -XGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

GenEmbl.*

| 1: | gb:ba.* |
|-----------------------|----------|
| 2: <td>gb:htg.*</td> | gb:htg.* |
| 3: <td>gb:in.*</td> | gb:in.* |
| 4: <td>gb:om.*</td> | gb:om.* |
| 5: <td>gb:ov.*</td> | gb:ov.* |
| 6: <td>gb:pat.*</td> | gb:pat.* |
| 7: <td>gb:ph.*</td> | gb:ph.* |
| 8: <td>gb:pi.*</td> | gb:pi.* |
| 9: <td>gb:pr.*</td> | gb:pr.* |
| 10: <td>gb:ro.*</td> | gb:ro.* |
| 11: <td>gb:sts.*</td> | gb:sts.* |
| 12: <td>gb:sv.*</td> | gb:sv.* |
| 13: <td>gb:un.*</td> | gb:un.* |
| 14: <td>gb:vi.*</td> | gb:vi.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 705 | 100.0 | 2472 | 6 | AX799092 Sequence |
| 2 | 705 | 100.0 | 2472 | 9 | HUMB5A |
| 3 | 226 | 32.1 | 2118 | 9 | BT007100 Homo sapi |
| 4 | 226 | 32.1 | 2118 | 12 | BT008132 Synthetic |

| 5 | 226 | 32.1 | 2329 | 6 | CQ834708 |
|----|-----|------|------|----|-----------|
| 6 | 226 | 32.1 | 2339 | 6 | CQ834706 |
| 7 | 226 | 32.1 | 2435 | 6 | CQ834710 |
| 8 | 226 | 32.1 | 2450 | 9 | BC048348 |
| 9 | 226 | 32.1 | 2458 | 6 | CQ834714 |
| 10 | 226 | 32.1 | 2472 | 6 | CQ727867 |
| 11 | 226 | 32.1 | 2516 | 9 | BC020965 |
| 12 | 226 | 32.1 | 2760 | 6 | CQ492634 |
| 13 | 220 | 31.2 | 2898 | 9 | HSB809254 |
| 14 | 192 | 27.2 | 1796 | 9 | AK130548 |
| 15 | 186 | 26.4 | 2293 | 9 | HUMTMC |
| 16 | 155 | 22.0 | 1094 | 5 | CR386917 |
| 17 | 155 | 22.0 | 2567 | 5 | BC083234 |
| 18 | 155 | 22.0 | 2608 | 5 | BC046072 |
| 19 | 152 | 21.6 | 2735 | 10 | BC037612 |
| 20 | 152 | 21.6 | 3094 | 6 | AX799090 |
| 21 | 152 | 21.6 | 3094 | 10 | MUSTRPR |
| 22 | 146 | 20.7 | 440 | 6 | AX351312 |
| 23 | 145 | 20.6 | 2481 | 5 | BC087313 |
| 24 | 144 | 20.4 | 2470 | 6 | CQ834712 |
| 25 | 144 | 20.4 | 3666 | 10 | AK129027 |
| 26 | 140 | 19.9 | 455 | 6 | AX341209 |
| 27 | 128 | 18.2 | 487 | 6 | BD229538 |
| 28 | 102 | 14.5 | 414 | 6 | AX884408 |
| 29 | 102 | 14.5 | 414 | 6 | BD024018 |
| 30 | 102 | 14.5 | 487 | 6 | AR273912 |
| 31 | 102 | 14.5 | 487 | 6 | AR277493 |
| 32 | 102 | 14.5 | 487 | 6 | AR407768 |
| 33 | 102 | 14.5 | 487 | 6 | AR441618 |
| 34 | 102 | 14.5 | 487 | 6 | AX368945 |
| 35 | 101 | 14.3 | 307 | 6 | AR272646 |
| 36 | 101 | 14.3 | 307 | 6 | AR273528 |
| 37 | 101 | 14.3 | 307 | 6 | AR276227 |
| 38 | 101 | 14.3 | 307 | 6 | AR277109 |
| 39 | 101 | 14.3 | 307 | 6 | AR406502 |
| 40 | 101 | 14.3 | 307 | 6 | AR407384 |
| 41 | 101 | 14.3 | 307 | 6 | AR440352 |
| 42 | 101 | 14.3 | 307 | 6 | AR441234 |
| 43 | 101 | 14.3 | 307 | 6 | AR472510 |
| 44 | 101 | 14.3 | 307 | 6 | AX062748 |
| 45 | 101 | 14.3 | 307 | 6 | AX367665 |

ALIGNMENTS

| RESULT 1 | AX799092 | Sequence 11 from Patent WO03054008. | 2472 bp | mrna | linear | PAT 08-OCT-2003 |
|------------|--|-------------------------------------|---------|------|--------|-----------------|
| AX799092 | Sequence 11 from Patent WO03054008. | | | | | |
| LOCUS | AX799092 | | | | | |
| DEFINITION | Sequence 11 from Patent WO03054008. | | | | | |
| ACCESSION | AX799092 | | | | | |
| VERSION | AX799092.1 | GI:37605063 | | | | |
| KEYWORDS | | | | | | |
| SOURCE | Human sapiens (human) | | | | | |
| ORGANISM | Human sapiens | | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | |
| AUTHORS | Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| TITLE | Perreault, C. and McBride, K. | | | | | |
| JOURNAL | Mammalian SIMP protein, gene sequence and uses thereof in cancer therapy | | | | | |
| FEATURES | Patent: WO 03054008-A 11 03-JUL-2003; | | | | | |
| source | Compatisgene Inc. (CA) | | | | | |
| | Location/Qualifiers | | | | | |
| | 1. .2472 | | | | | |
| | /organism="Homo sapiens" | | | | | |
| | /mol_type="mrna" | | | | | |
| | /db_xref="taxon:9606" | | | | | |
| | /note="ITM1, accession No NM-002219" | | | | | |

ORIGIN

Alignment Scores:
Pred. No.: 0
Score: 705.00
Length: 2472
Matches: 705

| | | | |
|------------------------|---------|---------------|---|
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DB: | 6 | Gaps: | 0 |

US-10-028-384-12 (1-705) x AX799092 (1-2472)

| | | | |
|----|------|---|------|
| QY | 1 | MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrIleuLeuLysLeu | 20 |
| DB | 107 | ATGACTAAGATTGGATTTTTCGGATTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT | 166 |
| QY | 21 | LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg | 40 |
| DB | 167 | CTCATCTGTCAATGGCTGCTGATATTATCTTCTCCACTCGTCTGTTGCTGCTCGAGA | 226 |
| QY | 41 | PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu | 60 |
| DB | 227 | TTTGAAGTGTATTCCATGAGTTTGATCCGTACTTTAATATCGGACTACCAAGTTCCTG | 286 |
| QY | 61 | AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTryptyrProLeu | 80 |
| DB | 287 | GCTGAGGAGGGTTTATAAATTCCTAACTGGTTTGATGACCGAGCCTGGTACCCTTTG | 346 |
| QY | 81 | GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr | 100 |
| DB | 347 | GGACGAATCATGGAGGAAACAAATTACCCAGGTTTAATGATCATCCTGCTGTCGAATCTAC | 406 |
| QY | 101 | HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla | 120 |
| DB | 407 | CATGTACTCCATTTTTCACATCATCACCATCGCAATTCGGAAATGCTGTGTGTTCTCGGCC | 466 |
| QY | 121 | ProLeuPheSerSerPheThrSerIleValThrTyrIleuLeuThrLysGluLeuLysAsp | 140 |
| DB | 467 | CCTCTCTTCTCTCTCTCCATCCATCGTCACGTACGCTCTCTACCAAGAGCTCAAGGAT | 526 |
| QY | 141 | AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg | 160 |
| DB | 527 | GCAGGGGCTGGGCTTCTTCTGCTGCTGCCATGATGCTAGTTGCTCTGGATATATCTCCGA | 586 |
| QY | 161 | SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr | 180 |
| DB | 587 | TCGTGGCTGGCTCTCTATGATAATGAAGGAGATTGCCATCTTTGTGATGCTACTACCTAC | 646 |
| QY | 181 | TyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaLeu | 200 |
| DB | 647 | TACATGTGGATCAAGGCAGTAAAGACTGGTTCCATCTGTGGGCAGCTAAGTGTGCCCTT | 706 |
| QY | 201 | AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuIleAsnLeuIlePro | 220 |
| DB | 707 | GCTTATTCTACATGGTCTCGTCAATGGGAGGTTATGTGTCTGATCAACTTAATTCCT | 766 |
| QY | 221 | LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr | 240 |
| DB | 767 | CTCCACGCTCTGCTGATGCTCACAGGCGGTTTCTTCACGGATCTATGTGGCCTAC | 826 |
| QY | 241 | CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln | 260 |
| DB | 827 | TGTACTGTTTACTGCTGGGTACTATACTTTCTAGGCAGATCTCTTTGTGGGTTTCCAG | 886 |
| QY | 261 | ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis | 280 |
| DB | 887 | CCTGTCTTTCATCAGAGCACATGGCAGGGTTTGGGGTCTTTGGTCTCTGCCAGATCCAT | 946 |
| QY | 281 | AlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGlnValLeuPheArg | 300 |
| DB | 947 | GCCTTTGGGATCTACCTCGGCAGCAGATTGAATCCACAAATTTGAAGTTCCTTTCCGG | 1006 |
| QY | 301 | SerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThr | 320 |
| DB | 1007 | AGCGTCATCTCTCTGGTAGGCTTTGTCTCTCTCACGTGGGAGCTCTCTCTCATGTGTACA | 1066 |
| QY | 321 | GlyValIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys | 340 |

| | | |
|------|---|------|
| 1067 | GGAAAAATATCTCCTCGGACGGGGCGTTTCTACTCACTGCTGGATCCCTCTTATGCTTAAG | 1121 |
| 341 | AAenAenIleProIleIleAlaSerValSerGluHisGlnProThrThrThrIlePheSerSerTyr | 360 |
| 1127 | AAACAATCCCCATCATTTGCTGTGTGAGCATCAGCCACAACTGGTCTCTCATAC | 1186 |
| 361 | TyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer | 380 |
| 1187 | TATTTTGACTGGAGCTTCCTCGTCTTCAATGTTTCAGTTGGCTCTATTACTGGCTTATGC | 1246 |
| 381 | AAenLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSer | 400 |
| 1247 | AACTGTCTGATGCGCCGGATTTTATTCATCATGATAGTGTGACCAAGCATGATCTTTTCA | 1306 |
| 401 | AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle | 420 |
| 1307 | GCTGTATAGGTGGTCTAATGCTAGTGTGGCACTCTTATGAGCATTCCTCTCTGGCATT | 1366 |
| 421 | GlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLys | 440 |
| 1367 | GGAGTCTCCCAAGTGTCTCCACATACATGAGAATCTGGACATAGTGGCCGACAGCAAG | 1426 |
| 441 | LysSerIleLysGlnGlnAspSerThrTyrProIleLysIleGluValAlaSerGlyMet | 460 |
| 1427 | AAAGCAGAGAAGCAACAGGATTCACCTACCTATTAGATTGAATGGCAAGTGGGATG | 1486 |
| 461 | IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTyrValThrSer | 480 |
| 1487 | ATACTGTGTCATGGGTTTCTTTCTCATCACCTCACCTTTCATTCAACCTGGGTGACCACT | 1546 |
| 481 | GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlyIleArgIle | 500 |
| 1547 | GAGGCTACTCTTCTCGTCCATTGACTATTCTGCGCGTGGTGGGATGGCAGTAGGATC | 1606 |
| 501 | IlePheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAenThrProGluAspAla | 520 |
| 1607 | ATATTGATGACTTCCAGAGAAGATATTATGGCTGTGTCTCAATACTATCCAGAGGATGCG | 1666 |
| 521 | LysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle | 540 |
| 1667 | AAAGTCATGCTCTGTGGGATTATGGCTATCAGATTACAGCTATGGCAAAACCGAACAAAT | 1726 |
| 541 | LeuValAspAenAenThrTrpAenAenThrHisIleSerArgValGlyGlnAlaMetAla | 560 |
| 1727 | TTATGTGGACAATAACATCGAAATAATACCCATATTTCTCGAGTAGGGCAGCGCAATGCGC | 1786 |
| 561 | SerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal | 580 |
| 1787 | TCCACAGAGGAANAAGCTATAGATCATGAGGAGGCTCGATGCAGCTATGCTCGTCTGC | 1846 |
| 581 | IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAenLysPheLeuTrpMetVal | 600 |
| 1847 | ATTTTGTGAGGCGCTCACTGGGTATCTCTGTGATGATATCAACAAGATTTCTTTGGATGGTC | 1906 |
| 601 | ArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThrPro | 620 |
| 1907 | CGATTGGAGGGACACAGATACAGCAAAACATATCAGGAGATGACTATTACTCTCA | 1966 |
| 621 | ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr | 640 |
| 1967 | ACTGGGAGATTCCCGTGTGGACCGTAGAGTTCTCCAGTGTGCTCACTGCTCATGTATC | 2026 |
| 641 | LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe | 660 |
| 2027 | AAATGTGTACTATCGCTTTGGCAGGTTTACACAGAAGCAAGCGTCTCTCCAGGCTTT | 2086 |
| 661 | AspArgValArgAenAlaGluIleGlyAenLysAspPheGluLeuAspValLeuGluGlu | 680 |
| 2087 | GACCGTGCAGAAATGCTGAGATTGGGAATAAAGACTTTTGAGCTTGATGTCTCTGGAGAA | 2146 |
| 681 | GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValLysAspLeuAspAsnArg | 700 |
| 2147 | GCCTATACACAGAACTTGGCTGGTCAAGATATACAGGTAAAGGACTCGATTAATCGA | 2206 |

```

Qy 701 GlyLeuSerArgThr 705
Db 2207 GCGTGTGCAAGGACA 2221

RESULT 2
HUMBSA
LOCUS
DEFINITION Human putative transmembrane protein precursor (B5) mRNA, complete
cds.
ACCESSION L38961
VERSION L38961.1 GI:624703
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 2472)
AUTHORS Lissy,N.A., Bellacosa,A., Sonoda,G., Miller,P.D., Jhanwar,S.C. and
Testa,J.R.
TITLE Isolation, characterization, and mapping to human chromosome
11q24-25 of a cDNA encoding a highly conserved putative
transmembrane protein, TMC
JOURNAL Biochim. Biophys. Acta 1306 (2-3), 137-141 (1996)
MEDLINE 96221283
PUBMED 8634329
FEATURES
source
1..2472
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2472
/gene="B5"
1..106
/gene="B5"
107..2224
/gene="B5"
107..2224
/gene="B5"
note="putative transmembrane protein precursor"
/codon_start=1
/protein_id="AAB05994.1"
/db_xref="GI:624704"

sig_peptide
mat_peptide
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

/note="encodes putative transmembrane domain"
1013..1060
/gene="B5"
/note="encodes putative transmembrane domain"
1199..1246
/gene="B5"
/note="encodes putative transmembrane domain"
1310..1357
/gene="B5"
/note="encodes putative transmembrane domain"
1475..1522
/gene="B5"
/note="encodes putative transmembrane domain"
2223..2472
/gene="B5"
2455
/gene="B5"

3'UTR
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2472
Score: 705.00 Matches: 705
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-10-028-384-12 (1-705) x HUMBSA (1-2472)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 107 ATGACTAAGTTGGATTTTGGCAATTTCTCTATGAGAAGCAGGACACATTTTGAAGCTT 166
Qy 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 167 CTCATCTCTCAATGGCTGCTGCTATTTATCTCTCCACTCGTCTGTTGCTGCTCTGAGA 226
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 227 TTGAAAGTGTATTCATGATGATTTGATCCGTACTTTAATATATCGGACTACCAAGGTTCTTG 286
Qy 61 AlaGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrTyrProLeu 80
Db 287 GCTGAGGAGGGGTTTATATAATCCATAACTGGTTTATGACCGAGCGCTGTTACCTTTG 346
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaIleTyr 100
Db 347 GGACGAATCATTTGGAGAACAAATTTACCCAGGTTTAAATGATCACCCTCTCTCTCAATCTAC 406
Qy 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 407 CATGTACTCCATTTTTCACATCACCACATTCGCAATTCGGAATGCTGTGTGTTCTTGCC 466
Qy 121 ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysAsp 140
Db 467 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 526
Qy 141 AlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArg 160
Db 527 GCAGGGGCTGGGCTTCTTGTCTGCTGCCATGATTTGCTGTAGTTCTCTGTATATATCTCCGA 586
Qy 161 SerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyr 180
Db 587 TCTGTGGCTGGCTCTCTATGATAAAGAGCTGGTTCCATCTCTTGGCAGCTAAGTGTGCCCTT 646
Qy 181 TyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeu 200
Db 647 TACATGTGGATCAAGGACGCTAAAGACTGGTTCCATCTCTTGGCAGCTAAGTGTGCCCTT 706
Qy 201 AlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuLeuPro 220
Db 707 GCTTATTCTACATGGTCTCTGTCATGGGAGGTTATGTGTTCTCTGATCAACTTAATCTCT 766

```

QY 221 LeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyr 240
 DB 767 CTCACGTCCTCGTGCTGATGCTCACAGGCGGTTCTCTCACCGGATCTATGTCGCTAC 826
 QY 241 CysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln 260
 DB 827 TGTACTGTCTACTGCTCGGTGCTATATCTTCTAGCAGATCTCTCTGTGGGTTCAG 886
 QY 261 ProValLeuSerSerGluHisMetAlaGlyPheGlyValPheGlyLeuCysGlnIleHis 280
 DB 887 CCTGTCCTTTCATCAGACGACATGGCAGGCTTTGGGGTCTTTGGTCTCTGCCAGATCCAT 946
 QY 281 AlaPheValAspTyrLeuArgSerLeuAsnProGlnGlnPheGluValLeuPheArg 300
 DB 947 GCCTTTGTGGATTACCTCGGCACGAAATGGAATCCCAACAATTTGAAGTTCTTTTCCGG 1006
 QY 301 SerValIleSerLeuValGlyPheValLeuThrValGlyValAlaLeuMetLeuThr 320
 DB 1007 AGCGTCATCTCTCTGGTAGCTTGTCTCTCACCGTGGAGCTCTCTCTCATGCTGACA 1066
 QY 321 GlyIleSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaLys 340
 DB 1067 GGAATAATATCTCCCTCGAGCGGGCGTTCTACTCACTCTGCTGATCCCTCTTATGCTAAG 1126
 QY 341 AsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerTyr 360
 DB 1127 AACACATCCCATCAATGCTTCTGTCTGTGATCATCAGCCCAACCTGGTCTCTCATAC 1186
 QY 361 TyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheSer 380
 DB 1187 TATTTTACCTCGACGCTCTCTGCTTCTCATGTTTCCAGTTGGCTCTTATTTACTGCTTACG 1246
 QY 381 AsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPheSer 400
 DB 1247 AACCTGCTCATGTCGCGGATTTTATCATCATGATGTTGTGACCATGATGACTTTTCA 1306
 QY 401 AlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyIle 420
 DB 1307 GCTGTATGTTGGTCTCAATGCTAGTGTGGACCTGTTATGAGCAATCTCTCTGGCAT 1366
 QY 421 GlyValSerGlnValLeuSerThrTyrMetLeuAsnLeuAspIleSerArgProAspLys 440
 DB 1367 GGAGTCTCCAGGTGCTGTCACATACATGAAGATCTGGACATTAAGTCCGCCACAGCAAG 1426
 QY 441 LysSerIleGlnGlnInAspSerThrTyrProIleIleGluValAlaSerGlyMet 460
 DB 1427 AAGAGCAAGAACCAACAGGATTCACCTACCTATTAAGATTGAAGTGGCAAGTGGGATG 1486
 QY 461 IleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThrSer 480
 DB 1487 ATACTGGTCACTGCTTCTTCTCATCATCACCTACCTTCACTCACTGGTGGTACCAAGT 1546
 QY 481 GluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIle 500
 DB 1547 GAGGCTACTCTCTCCGTCATTTGTAATCTATCTGCGCGTGGTGGGATGGCAGTAGGATC 1606
 QY 501 IlePheAspAspPheArgGluAlaTyrTyrLeuArgHisAsnThrProGluAspAla 520
 DB 1607 ATATTTGATGACTTCCGAGAAGCATATATTGGCTTCGTCAATACTCCAGAGGATGCG 1666
 QY 521 LysValMetSerTrpTrpAspTyrGlyTyrGluIleThrAlaMetAlaAsnArgThrIle 540
 DB 1667 AAGGTCACTGCTGCTGGGATTAAGGCTATCATGATTTACAGCTATGGCAACCCGACAAAT 1726
 QY 541 LeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla 560
 DB 1727 TTAGTGACATAACATACATGAATAATACCATATTTCTCGAGTAGGCGACGCAATGGCG 1786
 QY 561 SerThrGluGlyLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuVal 580
 DB 1787 TCCACAGAGGAAAAGGCTATGAGATCATGAGGAGCTCATGCTAGCTATGCTGCTGTC 1846
 QY 581 IlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrpMetVal 600

DB 1847 ATTTTGGAGGCTCACTGGGTATTCTCTGATGATCAACAAGTTCTTTGGATGGTC 1906
 QY 601 ArgIleGlyGlySerThrAspThrGlyTyrHisIleLeuGluAsnAspTyrTyrThrPro 620
 DB 1907 CGGATTGGAGGAGCACAGATACAGGCAACATATCAGGAGATGACTATTATCTTCCA 1966
 QY 621 ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr 640
 DB 1967 ACTGGGAGTTCGCTGTCGACCGTGAAGTTCTCAGTGTCTCACTGCTCACTGATAC 2026
 QY 641 LysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGlyPhe 660
 DB 2027 AGATGTGTACTATCTGCTTTGGACAGGTTTACAGAACCAAGCGTCTCCAGGCTTT 2086
 QY 661 AspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu 680
 DB 2087 GACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGTTCTCTGGAGAA 2146
 QY 681 GlyTyrThrThrGluHisTrpLeuValArgIleTyrLysValIleAspLeuAspAsnArg 700
 DB 2147 GGTATACACACAGACATTTGCTGTGAGTATACAGATACAGGTAAAGGACCTGGATATCGA 2206
 QY 701 GlyLeuSerArgThr 705
 DB 2207 GCGTTGTCAAGGACA 2221

RESULT 3
 BT007100 2118 bp mRNA linear PRI 13-MAY-2003
 LOCUS Homo sapiens integral membrane protein 1 mRNA, complete cds.
 DEFINITION
 ACCESSION BT007100
 VERSION BT007100.1 GI:30583038
 KEYWORDS FLI CDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2118)
 AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
 TITLES Cloning of human full-length CDSs in BD Creator(TM) System Donor vector
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2118)
 AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.
 TITLES Direct Submission
 JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
 COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orfclones>.
 FEATURES
 source location/Qualifiers
 1..2118
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GH00707X1.0"
 /clone_lib="BD Creator(TM) CDS Library derived from MGC collection"
 /lab_host="DH5alpha T1 resistant"
 /note="Vector: pDNR-Dual"
 1..2118
 CDS

| | | | | | | | | | |
|--|-----|---|-----|--------------------------|--|------------------------|--|---|--|
| /codon_start=1 | | /product="Integral membrane protein 1" | | /protein_id="AAP35764.1" | | /db_xref="GI:30583039" | | /translation="MTKFGRLSRYEKQDRLKLLILSMAVLFSFSTRIFAVLRPFESV LHEFPDNYRTRFLAEQGYKPHNWFDDRAWPLGRIIGGTIYPLGLMITSAAIYHV LHPFHTIDINVCVFLAPLSPSTFTIYVHLTKELADAGALAAAMIAVPGYISR SVAGSYDNEDGIALFCMLATYVMKIKAVTGSICWAAKALAYEVMYSSWGGYVFLNL IPLVLVLMLTRSHRIYVACTVYCLGILSLSQISFVQFQVPLSEHMAAFVGL COIFAFVILSKSNPQFEVLFPSVLSLVGFVLLTVGALLMUTKISLWTRGFYLL DPSYAKNIPILIASVSEHQPTWSSYFDLQILVFPFVGLYFCFNLSDARIFIIMY QTSYMFAGVAVMLVLAVPMVCLISGIVSOVLSTYMKNLISRPDKSKKKQDDSTY PIKNEAGVLMVAFELITYFTSTVTSSEYSSPSIVLSARGGSGSRIIFDDFREA YVLFPHNTPEDAKVSWMDYGYQITAMANETILVDNNWNTHTSRVQGMASTEBKA YEIMRELDVSVLVI FGLTGYSSDDINKPLMVRIGGSTDTGKHIKENDYTPTEGEP RVDRGSPVLLANLWYKWCYVRFQVYTEAKRPPGPRVNAEIGNKDFELDVLEAY TTEHVLVRIYKVKLDNRLSRT" | |
| ORIGIN | | | | | | | | | |
| Alignment Scores: | | | | | | | | | |
| Pred. No.: | | 1.58e-226 | | Length: | | 2118 | | | |
| Score: | | 226.00 | | Matches: | | 674 | | | |
| Percent Similarity: | | 98.25% | | Conservative: | | 0 | | | |
| Best Local Similarity: | | 98.25% | | Mismatches: | | 6 | | | |
| Query Match: | | 32.06% | | Indels: | | 12 | | | |
| DB: | | 9 | | Gaps: | | 0 | | | |
| US-10-028-384-12 (1-705) x EF007100 (1-2118) | | | | | | | | | |
| QY | 1 | MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu | 20 | | | | | | |
| DB | 1 | ATGACTAAGTTTGGGATTTTGGGATTTGCCTATGAGAGGAGGACACACTTTTGAAGCTT | 60 | | | | | | |
| QY | 21 | LeuLleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg | 40 | | | | | | |
| DB | 61 | CTCATTCCTCAATGGCTGCTGATATATCTCTCCATCTGCTGCTGCTGCTGCTGCTG | 120 | | | | | | |
| QY | 41 | PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu | 60 | | | | | | |
| DB | 121 | TTTGAAGTGTTTATCCATGAGTTTGTATCCGTTACTTTAATATCCGACTACCAAGTTCTG | 180 | | | | | | |
| QY | 61 | AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu | 80 | | | | | | |
| DB | 181 | GCTGAGGAGGGTTTATATAATCCATTAACCTGTTGTTGATGACCGAGCCTGGTACCTTTG | 240 | | | | | | |
| QY | 81 | GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr | 100 | | | | | | |
| DB | 241 | GGACGAATCATTTGGAGGAGCAATTTACCCAGGTTTAAATGATCACCCTCTGCTGCAATCTAC | 300 | | | | | | |
| QY | 101 | HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla | 120 | | | | | | |
| DB | 301 | CATGTACTCCATTTTCCCATCATCCATGACATTCGGAATGCTGCTGTTCTCTGGCC | 360 | | | | | | |
| QY | 121 | ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA | 140 | | | | | | |
| DB | 361 | CCT | 418 | | | | | | |
| QY | 140 | spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA | 160 | | | | | | |
| DB | 419 | ATGCAAGGGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 478 | | | | | | |
| QY | 160 | rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT | 180 | | | | | | |
| DB | 479 | GATCTGTGGCTGGCTCTCTATGATAATGAGGGATTTGCCATCTTTTGTGATGCTACTACCT | 538 | | | | | | |
| QY | 180 | YrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpTrpAlaAlaLysCysAlaL | 200 | | | | | | |
| DB | 539 | ACTACATGTGGATCAAGGAGCAATAAGACTGGTTCATCTGTTGGGAGCTAAAGTGTGCC | 598 | | | | | | |
| QY | 200 | euaLaryPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuLleP | 220 | | | | | | |
| DB | 599 | TTGCTTATTTCTACATGCTCTGCTCATGGGAGGTTATGTTCTCTGATCAACTTAATTC | 658 | | | | | | |

```

1735 CTGTCATTTTGGAGCGCTCACTGGGATTCCTCTGATGATATCAACAGTTCTTTGG 1794
QY MetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyr 618
DB 1795 ATGTCGCGGATTTGGAGGAGCAGATACAGGCAACATATCAAGGAGAACTACTATTAT 1854
QY 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
DB 1855 ACTCCACTGGGAGTTCCGTGTGGACCGTGAAGGTTCTCCAGTGTCTGCTCAACGCTC 1914
QY 639 MetTyrLysMetCysTyrTyrA:GpHeGlyGlnVal:ThrGluAlaLysArgProPro 658
DB 1915 ATGTACAAGATGTTACTACTATCGCTTTGGACAGGTTTACACAGAAGCAAGCGTCTCCA 1974
QY 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeu 678
DB 1975 GCGTTTGACCGTGTCCGAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGCTCTG 2034
QY 679 GluGlu 680
DB 2035 GAGGAA 2040

RESULT 4
BT008132
LOCUS 2118 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens integral membrane protein 1 mRNA,
partial cds.
ACCESSION BT008132
VERSION BT008132.1 GI:30585102
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 2118)
AUTHORS Kalinine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length cDNAs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2118)
AUTHORS Kalinine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This cDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each cDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The cDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
Location/Qualifiers
1..2118
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH0070711.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/note="vector: pDNR-Dual"
1..>2118
/codon_start=1
/note="Mutations: 2117:Stop->Leu"
/transl_table=1
/product="Homo sapiens integral membrane protein 1"
CDS

```

```

/protein_id="AAP36824.1"
/db_xref="GI:30585103"
/translation="MKFGFRLSEYKQDILLKLLLSMAAVLSFSFRLPAVLRFSV
LHEFDYFNRTTRFLAEFGYKFNHFDRAWTPFLGRITGGTIIYGLMITISAAITRV
LHFPHITDIDRVNCFVLAFLSFSTIIVTLTKELDAGALLAAAMIAVPGYISR
SVAGSYDNEGIAIFCMILTYTWMKAVTGSICWAACALAYFYVMVSMGGYVFLNL
IFLHVLMILGRFSHRIYVACTVYCLGTILSMQISFVGQFVLSSEHMAAFGLNL
COHAFDVLSKLNPOQFEPVLSVLGVFLVLLTVGALLMTGKISPMWGRFYSL
DPSAKNNIPILASVSEHPITWSSYVFDQLVFMFPGVLYYCFNSLSDARFIMY
GVTMYFSAVMVRLMLVLAPVACILSGISYQVLSLYMKLDIRDPDKSKSGDSTY
PIKNEVASGMILVNAFLITITPRTSTWTSERVSPLSARSGDGSRIIFDPREA
YELRHNTPEPAKWSMDYGYQTAMANTLIVDNTWNTTHISRVGQAMASTEKA
YEMRLDVSYVLVIFGLTGYSDDKNKLWVRIGGSTDGKHKENDYITPTGEF
RYDRGSPVLNCLMVKMYRFQGYTEAKRPFQFDRVNEIGNKDFELDLVEAY
TTEHLVRIYKVKLDNRLSRL"
ORIGIN
Alignment Scores:
Pred. No.: 1,58e-226 Length: 2118
Score: 226.00 Matches: 674
Percent Similarity: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.06% Indels: 12
DB: 12 Gaps: 0
US-10-028-384-12 (1-705) x BT008132 (1-2118)
QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
DB 1 ATGACTAAGTTTGGGATTTTTCGATTTGCTTATGAGAAGCAGGACACACTTTTGAAGCTT 60
QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 61 CTCATTCTGTCATATGGCTGCTGTATTATTCCTTCCACTGCTGTTGTTGTTGTTGTTG 120
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
DB 121 TTGAAAGTGTATCCATGAGTTCGATCCGTACTTAAATATCGGACTTACGAGTTCCTG 180
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspAlaTyrTyrProLeu 80
DB 181 GCTGAGGAGGGTTTATATATTCATTAACCTGTTGATGACCGAGCTGTTGTTGTTGTTG 240
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
DB 241 GGACGAATCATTTGGAGGAACAATTTTACCAGGTTTAAATGATCACTCTGCTGCAATCTAC 300
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
DB 301 CATGTAATCATTTTTCACATCACTACCATTCGGAATGCTGTGTGTTGTTGTTGTTG 360
QY 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysAla 140
DB 361 CCTCTCTTCTCTCTCTCTCCACAC-CATGCTCACGTACCA-CCTTACCAGAGAGCTCAAG 418
QY 140 sPaLagIvalagIleuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
DB 419 ATGACGAGGCTGGGCTTCTTGTGCTGCTGCAATGATGTTGTTAGTTCTTGATATATCTCC 478
QY 160 rgservAlaAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT 180
DB 479 GATCTGTGGCTGGCTCCTTATGATAATAGAGGATTTGCCATCTTTTGTGATGCTACTCACCT 538
QY 180 YrTyrMetTyrIleLysAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaL 200
DB 539 ACTACATGTGGATCAAGGAGTAAGACTGGTTCATCTGTTGGGAGCAGTAAGTGTGCCC 598
QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuLeuLeu 220
DB 599 TTGCTTATTTCACATGTTCTGTCATGGGAGGTATGTGTTCTCTGATCAACTTAATTC 658
QY 220 roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 240

```


Db 659 CTCTCCACGCTCCTCGTCTGATGCTCACAGGCGCGTTCTCTCTCACCGGATCTATGTGCGCT 718
 Qy 240 YrCysThrValTyrCysLeuGlyThrLeuSerArg-GlnIleSerPheValGlyPhe 259
 Db 719 ACTGACTGTTTACTGCTGGGCACTATCTCTCTAT-GCAGATCTCTCTTGTGGGTTTC 777
 Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuGlyGlnIle 279
 Db 778 CAGCGTCTCTTCTATCAGAGCACATGGCAGC-CTTTGGGGTCTTTGGTCTCTCGCAGAT 836
 Qy 279 ehislalpheValaspTyrLeuArgSerLysLeuasnProGlnClnPheGluValLeuPh 299
 Db 837 CCATGCCCTTGTGGGATACCTCGCGCAGCAAGTTGAATCCACAAATTTGAAGTCTTTT 896
 Qy 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe 319
 Db 897 CCGGAGCGTCATCTCTCTGGTAGGCTTTGTCTCTCACCGTGGGAGCTCTCTCATGCT 956
 Qy 319 urhGlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuaspProSerTyrAl 339
 Db 957 GACAGAAATATCTCTCTGGACCGGGCGGTTCTCTCTCGCTGGATCCCTCTTATGC 1016
 Qy 339 alysnAsnIleProIleAlaSerValSerGluHisGlnProThrThrTrpSerse 359
 Db 1017 TAAGAACAATCCCACT 1076
 Qy 359 rTyrTyrPheLeuLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
 Db 1077 ATACTATTGTACCTGGAGCT 1136
 Qy 379 eSerAsnLeuSerAspAlaArgIlePheIleMetTyrGlyValThrSerMetTyrPh 399
 Db 1137 TAGCAACCTGCTGATGCCCGGATTTTATCATCATGTATGTTGACCAAGCATGTACTT 1196
 Qy 399 eSerAlaValMetAlaGluMetLeuValLeuAlaProValMetSer-IleLeuSerC 419
 Db 1197 TCAGCTGAATGGTGGCTTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1255
 Qy 419 lyleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
 Db 1256 GCATTGGAGTCTCCAGGCTGCTCCACATACATAGAGATCTGGACATAGTCTCCAG 1315
 Qy 439 spLysIleSerLysGlnaspSerThrTyrProIleLysIle-GluValAlaSer 458
 Db 1316 ACAGAGAGCAAGAGCAAGAGATCCACCTACCTATTAAGAA-TGAAGTGGCAAGT 1374
 Qy 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpVal 478
 Db 1375 GGGATGATCTGGTCTATGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1434
 Qy 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
 Db 1435 ACAGTGGAGGCTACTCTCTCCGCTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1494
 Qy 499 ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
 Db 1495 AGGATCATATTGATGACTTCGAGAGCATATTATTGGCTTCGTCTATAATCTCCAGAG 1554
 Qy 519 AspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
 Db 1555 GATGCAAGGTCATCTCTGGGGATTTGGCTATCAGATCAGATCAGATCAGATCAGATCAG 1614
 Qy 539 ThrIleLeuValAspAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
 Db 1615 ACAATTTTATGGAGCAATAACACATAGGAATAATACCCATATTTCTCAGTAGGCGAGCA 1674
 Qy 559 MetAlaSerThrGluGluLysAlaTyrGluLeuMetArgGluLeuAspValSerTyrVal 578
 Db 1675 ATGGCGTCCACAGAGAAAAGCCCTATGAGATCATGAGGGAGCTCGATGTCAGCTATGTG 1734
 Qy 579 LeuValIlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrp 598
 Db 1735 CTGGTCATTTTGGAGGCGCTCAGTGGGTATTTCTCTGATGATATCAACAAGTTCTTTGG 1794

Qy 599 MetValargIleGlyGlySerThrAspThrGlyLysHisIleLysGluLeuAspTyrTyr 618
 Db 1795 ATGTCCGGATTTGGGGGAGCAGATACAGCAACATATCAAGGAGATGACTATAT 1854
 Qy 619 ThrProThrGlyGluPheArgValaspGluGlySerProValLeuLeuasnCysLeu 638
 Db 1855 ACTCAACTGGGAGTTCCGTGGACCGTGAAGGTCTCCAGTGTGCTCACTGCTC 1914
 Qy 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
 Db 1915 ATGTACAGATGTTTACTATCGCTTTGGACAGTTTACACAGAACCCAGGCTCTCCA 1974
 Qy 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeu 678
 Db 1975 GGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCG 2034
 Qy 679 GluGlu 680
 Db 2035 GAGGAA 2040
 RESULT 5
 CQ834708 2329 bp DNA linear PAT 29-JUL-2004
 LOCUS Sequence 579 from Patent WO2004058805.
 DEFINITION CQ834708
 ACCESSION CQ834708
 VERSION CQ834708.1 GI:50834245
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 Matsuda, A. and Yoneta, S.
 T cell activating gene
 TITLE Patent: WO.2004058805-A 579 15-JUL-2004;
 JOURNAL Asahi Kasei Pharma Corporation (JP)
 FEATURES
 source
 1. 2329
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 119..2080
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAH05569.1"
 /db_xref="GI:50834246"
 /translations="MTKFGFLRLSYEKQDTLLKLILSMAAVLSFSTRFLAVRFESV
 IHEDPYFNRYTRFLAEGFYKFNWFDRAWYPLGRIGGTIYPLMITSAIYHV
 LHPHITIDRNVCVFLAFLFSFTTIVTYHLTKELKADAGLLAAATAVVPGVISR
 SVAGSDNEGIALFCMLTYMYIYVAYCTVYCTILSNQISFVGFQFVLSSEMAA
 FGVLGCIHAFVDYLRSLNPOQFVLRSLVSLVGVLLITVGLLMLTGLISPTWG
 RPYLLDPSYAKNNIFLIASVSEHQPTWSSYFFDLQLLVFMFPVGLIYCFNSLSK
 IFIMYGVTSKNNIFLIASVSEHQPTWSSYFFDLQLLVFMFPVGLIYCFNSLSK
 QDSTYIKNEVASGMLYMAFFLITPTFTSTWTSSEYSSPISLSARGGDSKRIIF
 DDFREAYWLRHTPEDAKVMSWDYQITAMANTILVNNNTWNTTHISRVGQAMA
 STEBEKVRDEGDSVLIYFGLTGYSDDIINKFLMVRIGGSDTGTGKH;KENDY
 TPGEPRVDEGDSVLIYFGLTGYSDDIINKFLMVRIGGSDTGTGKH;KENDY
 VLEBAYTEHWLVRIYKVDLDRGLSRI"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,71e-226 Length: 2329
 Score: 226.00 Matches: 407
 Percent Similarity: 98.55% Conservativity: 0
 Best Local Similarity: 98.55% Mismatches: 3
 Query Match: 32.06% Indels: 6
 Ds: 6 Gaps: 0
 US-10-028-384-12 (1-705) x CQ834708 (1-2329)
 Qy 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValaspTyrLeuArgSerLysLeu 290

Db 773 TTTGGGCTTTTGGTCTCTGCCAGATCCATGCTTTTGGATTACCTGCGCAGCAAGTTG 832
Qy 291 AsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
Db 833 AATCCACACANATTTGAAGTTCTTTCCGGAGGCTCATCTCTCTGGTAGGCTTTGCTCT 892
Qy 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyIleSerProThrPheArgPhe 330
Db 893 CTCACCGTGGAGCTCTCTCATGCTGACAGGAAATATCTCCCTGGAGGGGGGTTTC 952
Qy 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIleProIleIleAlaSerValSe 350
Db 953 TACCC-GCTGCTGATCCCTCTATGCTAAGACACACATCCCATCATTTGCTCTGTGTC 1011
Qy 350 rGluHisGlnProThrThrTrpSerSerTyrPheAspLeuGlnLeuValPheMe 370
Db 1012 TGAGCATCAGCCCAACCTGGTCTCATATCTATTTTGACCTGACGCTCTCGTCTTCAT 1071
Qy 370 tPheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleI 390
Db 1072 GTTTCAGTTGGCTCTATTACTGCTTTAGCACTGCTGATGCCGGAATTTTATCAT 1131
Qy 390 eMetTyrGlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuVal 410
Db 1132 CATGTATGCTGTGACCAAGATGATCTTTTTCAGCTGTAATGGTGGCTTCTATGCTGT 1191
Qy 410 uAlaProValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrM 430
Db 1192 GGCACCTGTTATG-G-CATTCTCTGTCATTTGGAGTCTCCAGTGCTGTCACATACA 1250
Qy 430 etLysAsnLeuAspIleSerArgProAspLysSerLysLysGlnGlnAspSerThrT 450
Db 1251 TGAAGAATCTGGACATAGCTGCTCCAGCAAGAGAGCAAGAGCAACAGGATTCACCT 1310
Qy 450 yrProIleLysIle-GluValAlaSerGlyWeIleLeuValMetAlaPhePheLeuIle 469
Db 1311 ACCCTATTAGAA-TGAAGTGGCAAGTGGGATGATGATCTGGTCTATGCTTTCTTCATC 1369
Qy 470 ThrTyrThrPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSerIleVal 489
Db 1370 ACTTACACCTTTTCATCAACTGGTGGTGGACAGTGGGCTACTCTCTCCGTCACATGTA 1429
Qy 490 LeuSerAlaArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyr 509
Db 1430 CTATCTGCGCTGCTGGGATGGCAGTAGGATCATATTTTCAITGACTTCGAGAGACATAT 1489
Qy 510 TyrTrpLeuArgHisAsnThrProGluAspAlaLysValMetSerTrpTrpAspTyrGly 529
Db 1490 TATTGGCTTCGTATAATCTCCAGAGGATGCCAAGGTCATGCTCTGGTGGGATTTATGGC 1549
Qy 530 TyrGlnIleThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsn 549
Db 1550 TATCAGATTACAGCTATGGCAACCCGCAACATTTTAGTGACAATAACACATGGAATAAT 1609
Qy 550 ThrHisIleSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIle 569
Db 1610 ACCCATATTTCTCAGTAGGCGAGCAATGGCTCCACAGAGAAAGCCATATGAGATC 1669
Qy 570 MetArgGluLeuAspValSerTyrValLeuValIlePheGlyIleLeuThrGlyTyrSer 589
Db 1670 ATGAGGAGCTCGATGTCAGTATGCTGCTGCTCATTTTGGAGGCTCTCATCGGTATTC 1729
Qy 590 SerAspIleAenLysPheLeuTrpMetValArgIleGlySerThrAspThrGly 609
Db 1730 TCTGATGATATCAACAAGTTCTTTTGGATGCTCCGATTTGGAGGAGCAACATACAGGC 1789
Qy 610 LysHisIleLysGluAspAspTyrTyrThrProThrGlyGluPheArgValAspArgGlu 629
Db 1790 AAACATATCAAGGAGATGATATTAATCTCCAACTGGGAGTTCGCTGGACCGTGAA 1849
Qy 630 GlySerProValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGln 649
Db 1850 GGTTCCTCAGTGTCTCACTGCTCATGATACAGATGTTTACTATCTGCTTTGGACAG 1909

Qy 650 ValTyrThrGluAlaLysArgProGlyPheAspArgValArgAsnAlaGluIleGly 669
Db 1910 GTTTACACAGAGCCCAAGCGCTCTCCAGGCTTACCGGTGTCCGAAATGCTGAGATTGGG 1969
Qy 670 AsnLysAspPheGluLeuAspValLeuGluIle 680
Db 1970 AATAAGACTTTGAGCTTGATGTCTCTGGAGAA 2002
RESULT 6
Qy 68034706 2339 bp DNA linear PAT 29-JUL-2004
LOCUS Sequence 577 from Patent WO2004058805.
Qy 68034706
ACCESSION Q834706
VERSION Q834706.1 GI:50834243
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Matsuda, A. and Yoneta, S.
TITLE T cell activating gene
JOURNAL Patent: WO 2004058805-A 577-15-JUL-2004;
Asahi Kasei Pharma Corporation (JP)
FEATURES
source 1..2339
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
29..595
/note="unnamed protein product"
/codon_start=1
/protein_id="CAH05568.1"
/db_xref="GI:50834244"
/translation="MLTYYMWIKAVKTSICWAACALAYFYVWSWGQYVPLNLI
PLHVLVLMTRGRFSHRIVAYCTVYICITLISQISFVGPQVLYSEHMAAFGVGLC
OIHAFVDVLRSLKLPQDFEVLFRSVISLEKYLFGRGVSTRCPWLRLRTTSPSLLLCL
SISPPQPHITLTCSSSSSCFQVCGLVF"
ORIGIN
Alignment Scores:
Pred. No.: 1,726-226 Length: 2339
Score: 226.00 Matches: 306
Percent Similarity: 98.71% Conservative: 0
Best Local Similarity: 98.71% Mismatches: 2
Query Match: 32.06% Indels: 4
Gaps: 0
US-10-028-384-12 (1-705) x Q834706 (1-2339)
Qy 373 ValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleMetTyr 392
Db 765 GTTGGCTCTATTACTGCTTTAGCAACCTGCTCATGCCCGGATTTTATCATCATGTAT 824
Qy 393 GlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaPro 412
Db 825 GGTGTGACCAAGCATGATCTTTTACGTGTAATGTTGGCTCTAATGCTAGTGTGGCACCT 884
Qy 413 ValMetSer-IleLeuSerGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAs 432
Db 885 GTTATGTG-CATTCTCTGTCATTGGAGTCTCCAGGTGCTGCCACATACATAGAA 943
Qy 432 nleuAspIleSerArgProAspLysLysSerLysGlnGlnAspSerThrTyrProIle 452
Db 944 TCTGGACATTAAGTCTGTCAGACAGAGCAAGAGCAAGAGCAAGAGTTCACCTACCTAT 1003
Qy 452 elysIle-GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyr 472
Db 1004 TAAGAA-TCAAGTGCAAGTGGATGATGATGCTGGTCTGCTCTCTCATCACCTACA 1062
Qy 472 hrPheHisSerThrTrpValThrSerGluAlaTyrSerSerProSerIleValLeuSerA 492

```

Db      1063  CTTTTTCATCAACCTGGGTGACCGAGGCGCTACTCTTCTCCGTCACCTGTTACTATCTG 1122
Qy      492  laargGlycylAspGlySerAlaGlyIleIlePheAspAspPheArgGluAlaIleTyrTrpL 512
Db      1123  CCGTGGTGGGATGGCAGTAGGATCAATTTGATGACTTCCGAGAGACATATTATGGC 1182
Qy      512  euArgHisAsnThrProGluAspAlaIleValMetSerTrpTrpAspTyrGlyTyrGlnI 532
Db      1183  TTCGTCATAATACTCCAGAGGATGCGAAGGTCATGCTGGTGGGATATGCTATCAGA 1242
Qy      532  leThrAlaMetAlaIleAsnArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisI 552
Db      1243  TTACAGCTATGCAACCGAACAAATTTAGTGGCAATAACACATGGAAATAATACCCATA 1302
Qy      552  leSerArgValGlyGlnAlaMetAlaSerThrGluGluIleAlaIleMetArgG 572
Db      1303  TTTCTCGAGTAGGCGAGCAATGGCGTCCACAGAGGAAAAGCCTATGAGATCATGAGGG 1362
Qy      572  luLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspA 592
Db      1363  ACTCGATGTCAGCTATGTCGGTCAATTTTGGAGGCTCAGTGGGTATTCCTCTGATG 1422
Qy      592  spIleAsnLysPheLeuTrpMetValArgIleGlyGlySerThrAspThrGlyLysHisI 612
Db      1423  ATATCAACAAGTTCTTTGGATGGTCCGATTTGGAGGAGCACAGATACAGGCAACATA 1482
Qy      612  leIysGluAsnAspTyrTrpThrProThrGlyGluPheArgValAspArgGluGlySerP 632
Db      1483  TCAGAGAGANTGACTATTATCTCCACTGGGAGTTCCGTTGGAGCCGTGAGGTTCTC 1542
Qy      632  roValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrI 652
Db      1543  CAGTGTGCTCAACTGCCTCATGTACAAAGATGTGTACTATCGCTTTGGACAGGTTTACA 1602
Qy      652  hrGluAlaIleArgProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysA 672
Db      1603  CAGAAGCCAGGCTCTCCAGCTTTGACCGTGTCCGAATGCTGAGATGGGAATAAG 1662
Qy      672  spPheGluLeuAspValLeuGluGlu 680
Db      1663  ACTTTGAGCTTGATGTCCTCGAGGAA 1688

RESULT 7
LOCUS   CQ834710
DEFINITION Sequence 581 from Patent WO2004058805.
ACCESSION CQ834710
VERSION   CQ834710.1
KEYWORDS Homo sapiens (human)
SOURCE   Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Matsuda, A. and Yoneta, S.
TITLE T cell activating gene
JOURNAL Patent: WO 2004058805-A 581 15-JUL-2004;
Asahi Kasei Pharma Corporation (JP)
FEATURES
     Location/Qualifiers
     1..2435
     /organism="Homo sapiens"
     /mol_type="unassigned DNA"
     /db_xref="taxon:9606"
     73..2187
     /note="unnamed protein product"
     /codon_start=1
     /protein_id="CAH05570.1"
     /db_xref="GI:50834248"
     /translation="MTKFGELRLSYEQDILLKLLILSVAALVSFSTRFLFAVLRRESV
     IHEDPYFNRTTFLAEGFYKEHNFWDRAWYPLRIGGTIYFGLMTSAALYHV
     LHFFHTIDNVCVFLAPLPSFTTIVTLTKELKADAGALLAAVAVPGVIR
     SVAGSDNEGIAIFCMLITYMVIKAVTGSICWAAKCALAFYWSWGQYVFLNLI
     PLRVLMMLTRFSHRIVAYCTVCLGTTILSMQISFVGFPVLSSEHMAAGVFLCLC

```

CDS

```

OIHAFVDYLBKSLNPGQFEVLPRSVLSLVGPFVLLTYGALLMLTKGISPWTGRFYSLLD
PSYAKNNPIIASVSEHPTSSYIFDLQLVMEFVGLTYCFNSLDRKPIFLIMYG
VTSNYFAVMVRLWLVLAPVMCILSGVSVLSTYKKNLDSRPDKKKQDDSTYP
IKNEAVGMLVWAFELITYFHSWTSEAVSSPILVSARGGDSRIIFDDFREAY
YMLRNPEDAKVMWMDGYQITANARTILVDNNTWNTTHSRVQAMASTEKAY
EIMRLDVSIVLIFGLTGYSSDDINKFLMWRIIGSDTGTGKHKENDYIPTGEFR
VDRGSPVLNCLMYKCYVRFQVTEAKRPPGDFVRNABIGNKDPDLDELEYAT
TEHWLVRIYKVKLDNRGLSRT"
ORIGIN
Alignment Scores:
Pred. No.: 1,78e-226      .Length: 2435
Score: 226.00           Matches: 671
Percent Similarity: 97.67%      Conservative: 0
Best Local Similarity: 97.67%      Mismatches: 9
Query Match: 32.06%             Indels: 16
DB: 6                       Gaps: 0
US-10-028-384-12 (1-705) x CQ834710 (1-2435)

```

```

Qy      1  MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
Db      73  ATGACTAAGTTTGGATTTTTCGATTTGCTCTATGAGAAGCAGGACACACTTTTGAAGCTT 132
Qy      21  LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db      133  CTCATTCTGTCATCATGGTGTGTATTATCTCTTCCACTCGTCTGTTCTGTCGAGGA 192
Qy      41  PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db      193  TTTGAAAGTGTATTCATGAGTTTGTATGATCCGCTACTTTAAATATCGGACTACCAAGTTC 252
Qy      61  AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
Db      253  GCTGAGGAGGGGTTTATAAATTCATGCTGTTGTGACGAGCCCTGTTACCTTTG 312
Qy      81  GlyArgIleleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db      313  GGAGCAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATACACCTCTGCTGCAATCTAC 372
Qy      101  HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db      373  CATGACTCCATTTTTCACATCACCATCGACATTCGGAAATGCTGCTGTTCTCGGCC 432
Qy      121  ProLeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrLysGluLeuLysA 140
Db      433  CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 490
Qy      140  spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
Db      491  ATGAGGGGGCTGGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 550
Qy      160  rgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
Db      551  GATCTGTGGCTGGCTCTCTATGATAAAGAGGATTCGCATCTTTTGCATGCTACTCACCT 610
Qy      180  YrTyrMetTrpIleLysAlaValIleThrGlySerIleCysTrpAlaAlaLysCysAlaAl 200
Db      611  ACTACATGTGATCAAGGAGTAAAGACTGGTTCCATCTGTTGGGAGCTAGTAGTGCC 670
Qy      200  euAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheLeuIleAsnLeuLeuP 220
Db      671  TTG--CT--TTCTACATGGTCTCGTCATGGGAGGTTATGTTCTCTGATCACTTAATTC 727
Qy      220  roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaAla 240
Db      728  CTCTCCAGCTCTCTGTTGCTGATGCTCACAGGCCGTTTCTCTCACCGATCTATGTGGCT 787
Qy      240  YrCysThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPhe 259
Db      788  ACTGTACTGTATTACTCTGGGACATATACTTCTCTAT--GCAGATCTCTCTTTTGGTGTTC 846
Qy      260  GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279

```


Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 93 Row: p Column: 12
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22749414.

FEATURES
source

sources

```

1. .2450
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:5561615"
/tissue_type="Duodenum"
/clone_lib="NIH MGC_88"
/lab_host="DH103"
/note="Vector: pCMV-SPORT
1. .2450

```

CDs

```

/db_xref="WMIM:601134"
<1..2163
/gene="ITM1"
/codon_start=1
/product="ITM1 protein"
/protein_id="AAH48348.1"
/db_xref="GI:29477244"
/db_xref="locusID:3703"
/db_xref="WMIM:601134"

```

ORIGIN

| | | |
|------------------------|-----------|-----------------|
| Alignment Scores: | | |
| Pred. No.: | 1,79e-226 | length: 2450 |
| Score: | 226.00 | Matches: 674 |
| Percent Similarity: | 98.25% | Conservative: 0 |
| Best Local Similarity: | 98.25% | Mismatches: 6 |
| Query Match: | 32.06% | Indels: 12 |
| DB: | 9 | Gaps: 0 |

US-10-028-384-12 (1-705) x BC048348 (1-2450)

| | | | | | | | | | | | | | | | | | | | | | |
|-----|-----|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|------|-----|-----|
| 1 | Met | Thr | Lys | Phe | Gly | Phe | Leu | Arg | Leu | Ser | Tyr | Glu | Lys | Gln | Asp | Thr | Leu | Leu | Lys | Leu | 20 |
| 46 | AT | GACT | AAG | TTT | TGG | ATT | TTT | TGC | ATT | TTC | TAT | GAG | AAG | CAG | GAC | ACAC | ACT | TTT | GAG | CTT | 105 |
| 21 | Leu | Leu | Ser | Met | Ala | Ala | Val | Leu | Ser | Phe | Ser | Thr | Arg | Leu | Phe | Ala | Val | Leu | Arg | 40 | |
| 106 | CT | CA | TTCT | GTC | CA | TGG | CTG | TG | TAT | TA | CA | CTT | CTC | CA | CTG | CTG | TTT | GCT | GTCT | GAG | 165 |
| 41 | Phe | Glu | Ser | Val | Ile | His | Glu | Phe | Asp | Pro | Tyr | Phe | Asn | Tyr | Arg | Thr | Thr | Arg | Phe | Leu | 60 |
| 166 | TTT | GAA | AGT | GT | TAT | CC | ATG | AGT | TTG | TAT | CGC | TAC | TTT | AAT | TAT | CGG | ACT | ACC | AGG | TC | 225 |
| 61 | Ala | Glu | Glu | Gly | Phe | Tyr | Lys | Phe | His | Asn | Trp | Phe | Asp | Asp | Arg | Ala | Trp | Tyr | Pro | Leu | 80 |
| 226 | GCT | GAG | GAG | GGG | TTT | TAT | AA | TAT | CC | ATA | CTG | TTG | AT | GAC | CGC | CTG | TGG | TAC | CC | TTG | 285 |
| 81 | Gly | Arg | Ile | Leu | Gly | Thr | Ile | Tyr | Pro | Gly | Leu | Met | Ile | Thr | Ser | Ala | Ala | Ile | Trp | 100 | |

```

Db 1361 ACAAGAAGAGCAAGCAAGCAAGTATCCACCTACCTATTAAGAA-TGAAGTGGCAAGT 1419
Qy 459 GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal 478
Db 1420 GGGATGATACCTGGTCATGGCTTTCTTCTCATCACTACACCTTTTCATCACTGGGGT 1479
Qy 479 ThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
Db 1480 ACCAGTGGGGCTACTCTCTCCGTCCTATGTTACTATCTGCTGCGCTGGTGGGATGGCAGT 1539
Qy 499 ArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
Db 1540 AGGATCATATTGATGACTCCGAGAAGCATATTTATGGCTTCGTCATATATCTACGAG 1599
Qy 519 AspAlaIysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArg 538
Db 1600 GATCGAAGGTTCATGCTCTGCTGGGATTAATGGCTATCAGATTACAGTATGGCAACCGA 1659
Qy 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
Db 1660 ACAATTTAGTGGCAATATACACATGGAATAATACCATATTTCTCGATAGCGAGGA 1719
Qy 559 MetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal 578
Db 1720 ATGGCGTCCACAGAGGAAAGCCATATGAGATCATGAGGAGCTCGATGTCAGCTATGTG 1779
Qy 579 LeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnIysPheLeuTrp 598
Db 1780 CTGGTCATTTTGGAGCGCTCCTGGGTATTCCTCTGATGATATCAACAGTTCTTTGG 1839
Qy 599 MetValArgIleGlySerThrAspThrGlyGlyHisIleIysGluIleAsnAspTyrTyr 618
Db 1840 ATGGTCCGATTGGAGGAGCACAGATACAGGCAAAATATCAAGGAGATCACTATTAT 1899
Qy 619 ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu 638
Db 1900 ACTCCACTGGGAGTTCGGTGGACCGGTGAAGGTTCTCCAGTGTCTCACTGCCTC 1959
Qy 639 MetTyrIysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaIysArgProPro 658
Db 1960 ATGTACAAGATGTCTTACTATCGCTTGGACAGGTTTACACAGAACCAAGGCTCCTCA 2019
Qy 659 GlyPheAspArgValArgAsnAlaGluIleGlyAsnIysAspPheGluLeuAspValLeu 678
Db 2020 GCGTTTGACCTGTCCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTGATGCTCG 2079
Qy 679 GluGlu 680
Db 2080 GAGGAA 2085

RESULT 9
CO834714 2458 bp DNA linear PAT 29-JUL-2004
LOCUS
DEFINITION Sequence 585 from Patent WO2004058805.
ACCESSION CO834714
VERSION CO834714.1 GI:50834251
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. Matsuda, A. and Yoneta, S.
T cell activating gene
Patent: WO 2004058805-A 585 15-JUL-2004;
Asahi Kasei Pharma Corporation (JP)
LOCATION/Qualifiers
1. .2458
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
93. .2210
/feature="unnamed protein product"
CDS

```

```

/codon_start=1
/protein_id="CAH055572.1"
/db_xref="GI:50834252"
/translation="MTKEGFLSYEKQDTLLKLLILSMAAVLSFSFRLFAVLRFESV
LHFFITIDIRNVCFVLAFEGYKFNHFWEDRAWTPGLRIIGTIYFGLMITSAAIYRV
SVAGSYDNEGIAFCMLLTYVMKAVKATGICWAACKALAYFVWSWGYYFLNL
IFLHVLMLTGRFSHRIYVAYCTVCLGTILSMQISFVGFPVLSERHMAAFGVFL
CQHAFAVDYLRSLNPOQFEVLFPSVLSLVGLVLLTVGALLMTGKISPTWGFYSLL
DSYAKNIPITASVSEHPTWSSYFDFLOLLVFPFVGLYCFSLSDARFIFIY
GYTSVFSAVMRLMLVLPWCILSGIGNSQVLSYMKLDIRPDKSKKQODSY
PIKNSVASCMILLNAPFLITTYEHSWTSEYSPSLVLSARGDGRILIPDFEA
YWLKRNTPEDAKVMSWDYGYQTAMANRILVDNNTWNTHTHSRVQMASTESKA
YIMKELDVSVLVIFGGUTYSSDDINKFLMMVRIGSDTGHKIKENDIYYITPGEF
RYDRGSPVSLNCLMYKCYRFGQVTEAKRPFGRFDRVRNAEIGNKDFELDLVEAY
TTHEWLVRIVYKDLNDRGLSRT"
ORIGIN
Alignment Scores: 1.8e-226 Length: 2458
Pred. No.: 226.00 Matches: 674
Score: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.06% Indels: 12
DB: 6 Gaps: 0
US-10-028-384-12 (1-705) x CO834714 (1-2458)
Qy 1 MetThrIysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 93 ATGACTAAAGTTTGGATTTTTCGATTGTCTTATGAGAGCAGGACACACTTTTGAAGCTT 152
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 153 CTCATTCTGCTCAATGGCTGCTATTATCTTCTCCACTCTCTGTGTTGCTGCTCTGAGA 212
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
Db 213 TTGGAAGTGTATTCCATGAGTTTGCATCGTACTTAAATATCGGACTTACAGGTTCTCTG 272
Qy 61 AlaGluGluGlyPheTyrIysPheHisAsnTrpPheAspAspArgAlaTyrTyrProLeu 80
Db 273 GCTGAGAGGGGTTTATAAATCCATTAACCTGGTTGATGACCGAGCTGTGTACCTTTG 332
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 333 GGACGAATCATTTGGAGGAACAATTTACCAGGTTTATGATGATCCTCTGTGCAATCTAC 392
Qy 101 HisValIleHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 393 CATGTACTCCATTTTCCACATCACCATCGACATTCGGAATGCTGTGTGTTCTCTGCC 452
Qy 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrIysGluLeuLysA 140
Db 453 CCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 510
Qy 140 spAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
Db 511 ATGACGGGCTGGGCTTCTGCTGCTGCATCATTTGCTGTAGTTCTCTGGATATATCTCCC 570
Qy 160 rGserValIalaglySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
Db 571 GATCTGTGGCTGGCTCTCTATGATAATGAAGGATTTGCCATCTTTTGTGATGCTACTCACCT 630
Qy 180 YrTyrMetTrpIleLysAlaValIysThrGlySerIleCysTrpAlaAlaLysCysAlaL 200
Db 631 ACTACATGTGGATCAAGGCAGTAAAGACTGGTTCCATCTCTGTGTGGCAGCTAAGTGTGCC 690
Qy 200 eunlatYrPheTyrMetValSerSerTrpGlyTyrValPheIleLeuLeuLeuLeu 220
Db 691 TTGCTTATTTCATCATGTTCTGTCATGGGAGGTTATGTGTTCTCTGATCAACTTAATTC 750
Qy 220 rOleuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaL 240

```


Db 751 CTCTCCAGCTCTCTGCTGATGCTCACAGCGGTTTCTCTCACCGGATCTATGCGCT 810
 Qy 240 YrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 Db 811 ACTGTACTGTTTACTGCTGGGCACTATACCTTTCTAT-GCAGATCTCTCTTGTGGGTTTC 869
 Qy 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIle 279
 Db 870 CAGCGTGTCTTTCATCAGACACATGCGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGAT 928
 Qy 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPh 299
 Db 929 CCATGCCCTTTGTGATTAACCTGCGCAGCAAGTTGAATCCACAAATTTGAAGTCTTTT 988
 Qy 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLe 319
 Db 989 CCGAGCGTCTCTCTGTAGAGTTTGTCTTCTCACGTGGAGCTCTCTCATGCT 1048
 Qy 319 uThrGlyLysLeuSerProThrThrGlyArgPheTyrSerLeuLeuAspProSerTyrAl 339
 Db 1049 GACAGGAAATATCTCCCTGGACGGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGC 1108
 Qy 339 aLysAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
 Db 1109 TAAGACACATCCCATCATCTGTTCTGTCTGACATCAGCCCAACACCTGGTCCCTC 1168
 Qy 359 rTyrTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPh 379
 Db 1169 ATACTATTTGACCTGCGAGTCTCTGCTTCTCATGTTTCCAGTTGGCCTCTATTACTGCT 1228
 Qy 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh 399
 Db 1229 TAGCAACCTGCTGATCCCGGATTTTATCATCATGATGATGGGTGACCAAGCATGTACT 1288
 Qy 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG 419
 Db 1289 TTCAGCTGTAATGTGGCTGTAATGCTAGTGGTGGACCTGTTATGTC-CATTCTCTCTG 1347
 Qy 419 lYleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProA 439
 Db 1348 GCATTGGAGTCTCCAGGCTCTGCCACATACATGAGATCTGGACATAAGTCGTCACAG 1407
 Qy 439 sPlyLysSerLysGlnAspSerThrTyrProIleLysIle-GluValAlaLaser 458
 Db 1408 ACAAGAGACCAAGACACAGATTCACCTACCTACCTTTCATCACTTCACTTCACTTCTG 1466
 Qy 459 GlyMetIleValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpVal 478
 Db 1467 GGGATGATCTGTCATGGCTTTCTTCTCATCCTACCTTTCATCACTTCACTTCACTTCTG 1526
 Qy 479 ThrSerGluAlaTyrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer 498
 Db 1527 ACCAGTGAGCCCTACTCTTCTCGTCAATGTTACTATCTGCCGCTGGTGGGATGGCAGT 1586
 Qy 499 ArgIleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGlu 518
 Db 1587 AGGATCATATTCATGACTTCCGAGAAGCATATATTGCTGCTCATATATCTCCAGAG 1646
 Qy 519 AspAlaLysValMetSerTrpTyrAspTyrGlyTyrGlnIleThrAlaMetAlaSerArg 538
 Db 1647 GATCGAAGTCTATGCTCGTGGGATTAAGCTATCAGATTACAGCTATGCAACACCGA 1706
 Qy 539 ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla 558
 Db 1707 ACAATTTAGTGACAAATACACATGGAATATATACCATATTTCTCGATAGGAGGCA 1766
 Qy 559 MetAlaSerThrGluGlnLysAlaTyrGluIleMetArgGluLeuAspValSerTyrVal 578
 Db 1767 ATGGCTGCCACAGAGAAAGCCCTATGAGATCATGAGGAGCTCGATGCTCAGCTATGCTG 1826
 Qy 579 LeuValIlePheGlyLeuThrGlyTyrSerSerAspIleAsnLysPheLeuTrp 598

Db 1827 CTGGTCATTTTGGAGGCTCACTGGGTATTCTCTGATGATATCAACAAGTTTCTTTGG 1886
 Qy 599 MetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyr 618
 Db 1887 ATGGTCCGATTGGAGGAGCAGATACAGCAACATATCAAGGAGATCACTATTAT 1946
 Qy 619 ThrProThrGlyGluPheArgValAspArgGlyGlySerProValLeuLeuAsnCysLeu 638
 Db 1947 ACTCCAACTGGGAGTTCCGTGTGACCGTGAAGTTCTCCAGTCTGCTCAACTGCCTC 2006
 Qy 639 MetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPro 658
 Db 2007 ATGTACAAGATGTGTACTATCGTTTGGACAGGTTTACACAGAGCCAGGCTCTCCA 2066
 Qy 659 GlyPheAspArgValArgAsnIleGlyLysAsnLysAspPheGluLeuAspValLeu 678
 Db 2067 GGCCTTGGACCGTGTCCGAAATCTGAGATTGGAAATAAGACTTTGAGCTTGTATGCTCT 2126
 Qy 679 GluGlu 680
 Db 2127 GAGGAA 2132

RESULT 10
 CQ727867
 LOCUS CQ727867 2472 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 13801 from Patent WO02068579.
 ACCESSION CQ727867
 VERSION CQ727867.1 GI:42295093
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
 TITLE Kits, such as nucleic acid arrays, comprising a majority of
 humanexons or transcripts, for detecting expression and other uses
 thereof
 JOURNAL Patent: WO 02068579-A 13801 06-SEP-2002;
 PE Corporation (NY) (US)
 FEATURES
 Location/Qualifiers
 source 1..2472
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.81e-226 Length: 2472
 Score: 226.00 Matches: 673
 Percent Similarity: 98.25% Conservative: 0
 Best Local Similarity: 98.25% Mismatches: 6
 Query Match: 32.06% Indels: 12
 DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x CQ727867 (1-2472)
 Qy 2 ThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLysLeuLeu 21
 Db 110 ACTAAGTTTGGATTTTTCGATTTCTCTATGAGAAGCAGGACACACTTTGAAGCTTCTC 169
 Qy 22 IleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPhe 41
 Db 170 ATTCTGTCAATGGTGTGTATATCTTCTCCACTCGTCTGTTTGTCTGCTGAGATT 229
 Qy 42 GluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeuAla 61
 Db 230 GAAAGTGTATCCATGATGTTGATCCGCTACTTTAATATCGACTACAGGTTCTCGCT 289
 Qy 62 GluGluGlyPheTyrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeuGly 81
 Db 290 GAGGAGGGGTTTTATAAATCCCATACTGGTTTGTATGACCGCCTGGTACCCTTTGGGA 349

| | | | |
|-----|------|--|------|
| Qy | 82 | ArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHis | 101 |
| Ddb | 350 | CGAATCATGTGGAGGAACAATTTACCCAGGTTTAATGATCACCTCTGCTGCAATCTACCAT | 409 |
| Qy | 102 | ValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaPro | 121 |
| Ddb | 410 | GTACTCCATTTTTCACCATCACATCCATCGACATTCGGAATGTCGTGTGTTCTCTGGCCCT | 469 |
| Qy | 122 | LeuPheSerSerPheThrSerIleValThrTyrLeuLeuThrTyrSgluLeuLysAspA | 141 |
| Ddb | 470 | CTCTTCTCTCTCTTCCACACCATCGTCACGTACGTACCACTTACCAAAAGAGCTCAAGGATG | 527 |
| Qy | 141 | IaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerArgS | 161 |
| Ddb | 528 | CAGGGGCTGGGCTTCTTGCTGCTGCCATGATGCTGTAGTTCTCTGGATATATCTCCCGAT | 587 |
| Qy | 161 | erValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetIleuLeuThrTyrT | 181 |
| Ddb | 588 | CTGTGGCTGGCTCCTATGATAATGAAGGGATTCGCATCTTTGCGATGACTCTACCTACT | 647 |
| Qy | 181 | YrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAlaLeuA | 201 |
| Ddb | 648 | ACATGTGGATCAAGCAGTAAGACTGGTTCATCTGTGTGGCAGCTAAGTGTGCCCTTG | 707 |
| Qy | 201 | IaTyrPheTyrMetValSerSerTrpGlyGlyTyrValPheLeuIleAsnLeuIleProL | 221 |
| Ddb | 708 | CTTATTTTACATGTGCTCGTCATGGGAGGTATGTGTTCTGTCATCAACTTAATTCCTC | 767 |
| Qy | 221 | euHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrC | 241 |
| Ddb | 768 | TCCAGTCTCTGTGCTGATGCTCACAGGCGGTTTCTCTACCGGATCATGTGGGCTACT | 827 |
| Qy | 241 | YsThrValTyrCysLeuGlyThrIleLeuSerArgGlnIleSerPheValGlyPheGln | 260 |
| Ddb | 828 | GTACTGTTTACTGCCTGGGCACATATACTTTCTATGCAGATCTCCTTTGTGGGTTTCCAG | 886 |
| Qy | 261 | ProValIleSerSerGluHisIleValAlaGlyPheGlyValPheGlyLeuCysGlnIleH | 280 |
| Ddb | 887 | CTGTCTCTTTCATCAGAGACATGGCAGCCTTTGGGGTCTTTGGTCTCTGCCACATCCA | 945 |
| Qy | 280 | sAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnInPheGluValLeuPheAr | 300 |
| Ddb | 946 | TGCTTTTGTGGAATACCTGGCAGCAAGTTGAATCCACAACAATTTGAAGTTCTTTTCGG | 1005 |
| Qy | 300 | gSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuTh | 320 |
| Ddb | 1006 | GAGCGTCATCTCTGATAGGCTTTGTCTCTTCACCGTGGGAGCTCTCTCTCATGCTGAC | 1065 |
| Qy | 320 | rGlyLysIleSerProTrpThrGlyArgPheTyrSerLeuLeuAspProSerTyrAlaL | 340 |
| Ddb | 1066 | AGGAAAAATATCTCCCTGGAGCGGGCGTTTCTACTCGCTGCTGGATCCCTCTTATGCTAA | 1125 |
| Qy | 340 | sAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTy | 360 |
| Ddb | 1126 | GAACAACATCCCATCATGCTCTGTGCTGAGCATCAGCCACCAACCTGGTCTCTCAFA | 1185 |
| Qy | 360 | rTyrPheAspLeuGlnLeuValPheMetPheProValGlyLeuTyrTyrCysPheS | 380 |
| Ddb | 1186 | CTATTTTGACCTGCAGCTCTCTGCTTCATGTTTCCAGTTGGCTCTTATTACTGCTTTAG | 1245 |
| Qy | 380 | rAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheS | 400 |
| Ddb | 1246 | CAACCTGTCTGATGCCGGATTTTATCATCATGTATGGTGCACCAAGATGACTCTTTC | 1305 |
| Qy | 400 | rAlaValMetValArgLeuMetLeuValLeuAlaProValMetSerIleLeuSerGlyI | 420 |
| Ddb | 1306 | AGCTGTAATGGTGGCTTATGTAGTGTGGACCTGTTATGTGCAATTCTCTCTGGCA | 1364 |
| Qy | 420 | IeGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspL | 440 |
| Ddb | 1365 | TTGAGTGCTCCAGTGCTGTCCACATACATGAAGAAATCTGGACATAAGTCTGTCAGACA | 1424 |
| Qy | 440 | vsIvsSerValLysGlnAspSerThrTyrProIleLysIleGluValAlaSerGly | 459 |

| | | | |
|---|----|---|------|
| 1425 | Db | AGAAGACGAAGCAACAGAGATTCACCACTACCCCTATTAAAG-TGAAGTGGCAAGTGGG | 1483 |
| 460 | Qy | MetIleuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpValThr | 479 |
| 1484 | Db | ATGATACTGGTCATGCGCTTCTTCTTCATCACCCTACACCTTCATTCACCTGGGTGACC | 1543 |
| 480 | Qy | SerGluAlaTyrSerSerProSerIleValIleuSerAlaArgGlyGlyAspGlySerArg | 499 |
| 1544 | Db | AGTGAGGCCCTACTCTCTCCGTCCTTGTACTATCTGCCCGTGGTGGGATGGCAGTAGG | 1603 |
| 500 | Qy | IleIlePheAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAsp | 519 |
| 1604 | Db | ATCATATTTGATGACTTCGAGAGCATATATTGGCTTCGTCAATATATCTCCAGAGGAT | 1663 |
| 520 | Qy | AlaIysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThr | 539 |
| 1664 | Db | CGCAAGGCTCATGTCCTGGTGGGATTATGGCTATCAGATTACAGTATGGCAACCGAACA | 1723 |
| 540 | Qy | IleIleValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMet | 559 |
| 1724 | Db | ATTTTGTGGACCAATACACATCGGAATATATCCCATATTTCTCAGTAGGCGAGGCATG | 1783 |
| 560 | Qy | AlaSerThrGluGluIysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeu | 579 |
| 1784 | Db | CGCTCCACAGAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGTCTG | 1843 |
| 580 | Qy | ValIlePheGlyIleuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMet | 599 |
| 1844 | Db | GTCAATTTTGGAGGCTCACTGGGTATTCCTCTGATGATATCAACAGTTTCTTTGGATG | 1903 |
| 600 | Qy | ValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrTyrThr | 619 |
| 1904 | Db | GTCCGGATTGGAGGCGACAGATACAGCGAAACATATCAAGGAGATGACTATTATACT | 1963 |
| 620 | Qy | ProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMet | 639 |
| 1964 | Db | CCAACTGGGAGTTCCGTGTGGACCGTGAAGTTCTCCAGTGTCTCTCACTGCTCATG | 2023 |
| 640 | Qy | TyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProGly | 659 |
| 2024 | Db | TACAGATGTGTACTATTCGTTTGACAGGTTTACACAGAGCCCAAGCGTCTCCAGGC | 2083 |
| 660 | Qy | PheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGlu | 679 |
| 2084 | Db | TTTGACCGTGTCCGAAATGCTGAGATTGGCAATAAAGACTTTTGAGCTTGATGTCTGGAG | 2143 |
| 680 | Qy | Glu 580 | |
| 2144 | Db | GAA 2146 | |
| RESULT 11 | | | |
| BC020965 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| Homo sapiens: integral membrane protein 1, mRNA (cDNA clone MGC:90424 | | | |
| IMAGE:3891543), complete cds. | | | |
| ACCESSION | | | |
| VERSION | | | |
| BC020965.2 GI:34190969 | | | |
| KEYWORDS | | | |
| MGC. | | | |
| SOURCE | | | |
| Homo sapiens (human) | | | |
| ORGANISM | | | |
| Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | |
| Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | |
| REFERENCE | | | |
| 1 (bases 1 to 2516) | | | |
| Klausner,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G., | | | |
| Schulster,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., | | | |
| Altchul,S.F., Zeeberg,B., Buettow,K.H., Schaefer,C.F., Shat,N.K., | | | |
| Hopkins,R.F., Jordan,H., Moore,T., Max,I., Wang,J., Hsieh,F., | | | |
| Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., | | | |
| Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., | | | |
| Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., | | | |
| Carrinson,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., | | | |
| Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J. | | | |

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, J.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2516)
Strausberg, R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 25, 2003 this sequence version replaced gi:18089244.
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcdepaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 14 Row: e Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 22749414.
Location/Qualifiers

1. .2516
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:9042 IMAGE:3891543"
/tissue_type="Lung, carcinoma, large cell undifferentiated."
/clone_lib="NIH_MGC_69"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .2516
/gene="ITM1"
/note="synonyms: TMC, MGC9042"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
93_2210
/gene="ITM1"
/codon_start=1
/product="integral membrane protein 1"
/protein_id="AAH20965.1"
/db_xref="GI:18089245"
/db_xref="LocusID:3703"
/db_xref="MIM:601134"
/translation="MTKFGRLSLVEKQDTLKLILSLNAVLSTFLRFLAVLFRESV
LHEFDYFNRITRFALAEFGYKFNHFDRAWYPLGRIGGTIYVPGMITSAAIYHV
LHFITIDIRNVCVFLAFPSFTITVYHLTKELKADGALAAAMAVVPGYISR
SVAGSDNEGIAIFCMLLTYNWKVAVTKTSCWAACKALAYFVWVSGGQVFLINL
IPLHVLMLTGRFSHRIYVAVCTVYCLGTILSMOISFVGFQPLVSSHHMAFGVGL
COIHAVDVLRSKLNAPQOEVLFRSVISLVGLVLTVALMLTKGISFWTRFYSLL
DPSYAKNPIIATSVSEHFTWSSYIFDLQLLVFVFPVGLYICFNSLSDARIIFIMY
GVTSMTFSVAVRLMLVLAFLVWISIGVGSQVLSYMKNDISRPDKSKKQDSTY
PIKNEVASGMILWMAFFLITFTFHTWTSEAYSPSIVLSARGDGSRIIFDPFREA
YYLRLHNTPEDAKVMWMDYGYQITAMANRITLVDNTWNNTNTHISRVCQAVASTEKA

FEATURES
source

gene

CDS

YEIMRELDVSVLVIFGGTGYSSDDINKFLMWVPIGGSTGTGKHKENDYVTPGEP
YDRSGSPVLLNCLMYKCYRFRPGVYFTEARPPGDFRVRNABIGNKDFELVDLEAY
TTEHVLVRIYKVLNDNLGSLRT"

ORIGIN

Alignment Scores:
Pred. No.: 1.83e-226 Length: 2516
Score: 226.00 Matches: 674
Percent Similarity: 98.25% Conservative: 0
Best Local Similarity: 98.25% Mismatches: 6
Query Match: 32.06% Indels: 12
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x BC020965 (1-2516)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
Db 93 ATGACTAAGTTTGGATTTCGATGTCCATGAGAGCAGGACACACTTTTGAAGTT 152
QY 21 LeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 153 CTCATTCTGCAATGGCTGCTGTAATTATCTTCTCCACTCGTCTGTTGCTCTCTGAGA 212
QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
Db 213 TTGAAAGTGTATCATGAGTTTGCATCGTACTTTAATATCGGACTACCAGSTTCCTG 272
QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrTyrProLeu 80
Db 273 GCTGAGGAGGGTTTTTAATAATCCATAACTGGTTTGATGACCGAGCTGTACCTTTG 332
QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 333 GGACCAANTCATGGAGGAAACAATTTACCAGGTTAATGATCAGCTCTGCTGCAATCTAC 392
QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
Db 393 CATGTACTCCATTTTTCACATCACCATCGACATTCGAAATGCTGTGTGTTCTCTGCC 452
QY 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA 140
Db 453 CCTCTCTCTCTCTCTCTCACCAC-CATCGTCACGTACCA-CCTTACCAAGAGCTCAAGG 510
QY 140 spAlaGlyValAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
Db 511 ATGACAGGGCTGGGCTCTTCTGCTGCATGATGTGTAGTTCCTGGATATATCTCCC 570
QY 160 rGSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrT 180
Db 571 GATCTGTGGCTGGCTCCTATGATGATGAAGGGATTGCCATCTTTTGCATGCTACTCACCT 630
QY 180 YrTyrMetTyrIleValAlaValLysThrGlySerIleCysTyrAlaAlaLysCysAlaL 200
Db 631 ACTACATGTGGATCAAGGCAGTAAAGACTGGTTCATCTGTGTCAGCTCAAGTGTGCC 690
QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIleP 220
Db 691 TTGCTTATTTCTACATGCTCTCGTCATGGGAGGTTATGTTCCTGATCACTAATTC 750
QY 220 roLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaT 240
Db 751 CTCTCCACGCTCTCTGCTGCTGATGCTACAGGCGCTTCTCTCACCAGGATCTAATGTG 810
QY 240 YrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
Db 811 ACTGTACTGTTTACTGCTGGGACATATATTTCTAT-GCAGATCTCTTTGTGGGTTTC 869
QY 260 GlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
Db 870 CAGCCTCTCTCTTTCATCAGAGCACATGGCAGC-CTTTGGGGTCTTTTGGTCTCTGCCAGAT 928
QY 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPh 299

| | | | |
|--|------|--|------|
| Db | 929 | CCATGCCCTTTGGGATTACCTGCGCAGCAAGTTGAATCCACAAATTTGAAGTCTTTT | 988 |
| Qy | 299 | eAgsSerValIleSerLeuValGlyPheValLeuThrValGlyAlaLeuLeuMetLe | 319 |
| Db | 989 | CCGGAGCGTCATCTCTGGAGGCTTTGTCTCTCACCCTGGAGCTCTCTCATGCT | 1048 |
| Qy | 319 | uThrGlyLysIleSerProTrrPThrGlyArgPheTy-SerLeuLeuAspProSerTyAl | 339 |
| Db | 1049 | GACAGAAAATATCTCCCTGAGCGGCGTTTCTACTCGCTGTGGATCCCTCTTATGC | 1108 |
| Qy | 339 | alyAsnLeuIleProIleIleAlaSerValSerGluHisGlnProThrTrpSerSe | 359 |
| Db | 1109 | TAAGAACACATCCCATCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1168 |
| Qy | 359 | rTyTyTyPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyTyTyCysPh | 379 |
| Db | 1169 | ATACTATTGTGACCTGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 1228 |
| Qy | 379 | eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyTyGlyValThrSerMetTyPh | 399 |
| Db | 1229 | TAGCAACCTGTCTGTATGCCGCGATTTTATCATCATGTATGTGTGTGTGTGTGTGT | 1288 |
| Qy | 399 | eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerG | 419 |
| Db | 1289 | TTGAGCTGTAAATGGTGGCTCTAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1347 |
| Qy | 419 | lyIleGlyValSerGlnValLeuSerThrTyMetLysAsnLeuAspIleSerArgProA | 439 |
| Db | 1348 | GCATTCGAGCTCTCCAGGT | 1407 |
| Qy | 439 | spLysLysSerLysGlnGlnAspSerThrTyTyProIleLysIle-GluValAlaSer | 458 |
| Db | 1408 | ACAAGAAGAGCAAGAACAGGATTCACCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCT | 1466 |
| Qy | 458 | GlyMetIleLeuValMetAlaPhePheLeuIleThrTyTyPheHisSerThrTrpVal | 478 |
| Db | 1467 | GGGATCATCTGTCTGTATGCTTTCTTCATCATCATCATCATCATCATCATCATCATCAT | 1526 |
| Qy | 478 | ThrSerGluAlaTySerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySer | 498 |
| Db | 1527 | ACCAGTGAGCGCTACTCTCTCTCGTCCATTTGACTATCTGCGCGGTGGGATGGCAGT | 1586 |
| Qy | 498 | ArgIleIlePheAspAspPheArgGluAlaTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy | 518 |
| Db | 1587 | AGGATCATATTGATGACTCCGAGAGCATATTATTGCTTCGTATTAATCTCATTAATCT | 1646 |
| Qy | 518 | AspAlaLysValMetSerTrpTrpAspTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy | 538 |
| Db | 1647 | GATCGAAGGTCTATGTCCTGGTGGGATTTATGCTATCATGATTTACAGCTATGGCAACCGA | 1706 |
| Qy | 538 | ThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAla | 558 |
| Db | 1707 | ACAAATTTAGTGACCAATACACATGATATATATATATATATATATATATATATATATAT | 1766 |
| Qy | 558 | MetAlaSerThrGluGluLysAlaTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy | 578 |
| Db | 1767 | ATGCGGTCCACAGAGAAAAGCCTATGATCATGAGGAGCTCGATGTCAGCTATGCTG | 1826 |
| Qy | 578 | LeuValIlePheGlyGlyLeuThrGlyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy | 598 |
| Db | 1827 | CTGGTCATTTTGGAGGCTCACTGGGTATTCCTCTGTATGATATCAACAAAGTTTCTTTGG | 1886 |
| Qy | 598 | MetValArgIleGlySerThrAspThrGlyIleHisIleLysGluAsnAspTyTyTyTy | 618 |
| Db | 1887 | ATGGTCGGGATTTGGAGGAGCAAGATACAGGCAACATATCAAGGAGATCACTATTAT | 1946 |
| Qy | 618 | ThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeu | 638 |
| Db | 1947 | ACTCAACTGGGAGTTCCGTGTGGACCGGGAAGGTTCTCCAGTGTCTCACTCACTGCCTC | 2006 |
| Qy | 638 | MetTyTyLysMetCysTy | 658 |
| Db | 2007 | ATGTACAGATGTGTACTGCTTTGGACAGGTTTACAGAGCCCAAGCGTCTCTCCA | 2066 |
| Qy | 659 | GlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeu | 678 |
| Db | 2067 | GGCTTTGACGCTGCGGAATGCTGAGATGGGAATAAGACTTTGAGCTTGAATGCTCTG | 2126 |
| Qy | 679 | GluGlu | 680 |
| Db | 2127 | GAGGAA | 2132 |
| RESULT 12 | | | |
| CO492624 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| ORIGIN | | | |
| Alignment Scores: | | | |
| Pred. No.: | | | |
| Score: | | | |
| Percent Similarity: | | | |
| Best Local Similarity: | | | |
| Query Match: | | | |
| DB: | | | |
| US-10-028-384-12 (1-705) x CO492624 (1-2760) | | | |
| Qy | 1 | MetThrTyPheGlyPheLeuArgLeuSerTyTyGlyGlnAspThrLeuLeuLysIleu | 20 |
| Db | 105 | ATGACTAAGTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTT | 164 |
| Qy | 21 | LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr | 40 |
| Db | 165 | CTCATCTGTCAATGGCTGCTGTATTATCTTCTCCACTCGTCTGTTTGTCTGCTGAG | 224 |
| Qy | 40 | gPheGluSerValIleHisGluPheAspProTyTyPheAsnTyTyArgThrArgPheIle | 60 |
| Db | 225 | ATTGAAAGTGTATCCATGAGTTTGGATTTTGGATTTTGGATTTTGGATTTTGGATTTT | 284 |
| Qy | 60 | uAlaGluGluGlyPheTy | 80 |
| Db | 285 | GGCTGAGAGGGGTTTATATAATCCATACCTGTTTGTATGACCGAGCTGTGACCTTT | 344 |
| Qy | 80 | uGlyArgIleIleGlyThrIleTy | 100 |
| Db | 345 | GGGAGCAATCATTTGGAGAACAAATTTACCCAGGTTTAAATGATCACCTCTGCTGCAATCTA | 404 |
| Qy | 100 | rHisValLeuHis-PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuA | 120 |
| Db | 405 | CCATGTACTCTCATTTTTCATCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCAT | 464 |
| Qy | 120 | laProLeuPheSerSerPheThrSer-IleValThrTyTyLeu-LeuThrTyTyGluLeuTy | 139 |
| Db | 465 | CCCT | 522 |
| Qy | 139 | sAspAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValProGlyTyTyIleSe | 159 |

Db 523 GGATCCAGGGGCTGGGCTTCTGCTGCGCCATGATTGCTGTAGTTCCTCGATATATCTC 582
 Qy 159 rArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuTh 179
 Db 583 CGGATCTGCTGGCTGCTCTATGATATGAAGGATTCCTATCTTTGGCATGCTACTCAC 642
 Qy 179 rTyrTyrMetTyrPheLeuValAlaValLysThrGlySerIleCysTyrPalaAlaLysCysAl 199
 Db 643 CTACTACATGTGGATCAAGGAGTAAGACCTGGTTCATCTGTGGGCACCTAAGTGTGC 702
 Qy 199 aLeuAlaTyrPheTyrMetValSerSerTyrGlyTyrValPheLeuLeuLeuIle 219
 Db 703 CCTTGCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTGTCTCGTATCACTAAT 762
 Qy 219 eProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAl 239
 Db 763 TCTCTCCAGTCTCTGGTGTGATGCTCACAGGCGGTTTCTCTCCCGATCTATGTGGC 822
 Qy 239 aTyrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyP 259
 Db 823 CTACTGTACTGTACTGCTGGGCACCTATATCTTCTAT-CCAGATCTCTCTTGTGGTT 881
 Qy 259 heGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGln 278
 Db 882 TCCAGCTGCTCTTTCATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAG 940
 Qy 279 IleHisAlaPheValAspTyrIleuArgSerIleuLeuAsnProGlnInPheGluValLeu 298
 Db 941 ATCCATGCTTGTGGATTAATCTGGCAGCAAGTTGAATCCACAAATTTGAATTCCT 1000
 Qy 299 PheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMet 318
 Db 1001 TTCGGAGGGTCACTCTCTGGTAGGCTTGTCTCTTCCACCGTGGAGGCTCTCTCCTG 1060
 Qy 319 LeuThrGlyLysIleSerProTyrThrGlyArgPheTyrSerLeuLeuAspPProSerTyr 338
 Db 1061 CTGACAGGAAAAATATCTCCCTGGACGGGGGCTTCTACTCGCTGCTGGATCCCTCTAT 1120
 Qy 339 AlaLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTyrSer 358
 Db 1121 GCTAAGAACATCCCATCATCTCTGCTGTGCTGAGCATCAGCCCAACACTGTCTCC 1180
 Qy 359 SerTyrTyrPheAspLeuLeuLeuValPheMetPheProValGlyLeuTyrTyrCys 378
 Db 1181 TCATACATATTTGACCTGCAGCTCTCTCTCTATGTTTCCAGTTGGCTCTATTACTGC 1240
 Qy 379 PheSerAsnLeuSerAspAlaArgPheIleIleMetTyrGlyValThrSerMetTyr 398
 Db 1241 TTTAGCAACCTCTGTATGCCGGATTTTATCATCATGATGATGATGATGATGATGATGAT 1300
 Qy 399 PheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSe 418
 Db 1301 TTTTTCAGCTGTAAATGGTGGTCTAATGCTAGTGTGGCAGCTGTTATGTG-CATTCTCTC 1359
 Qy 418 rGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgPr 438
 Db 1360 TGGCATGTAGTCTCCAGTCTGCTGCACATGATGATGATGATGATGATGATGATGATGAT 1419
 Qy 438 oAspLysLysSerLysGlnAspSerThrTyrProIleLysIle-GluValAlaAs 458
 Db 1420 AGACAAGAGACAGACAGACAGATTCACCTTACCTTACCTTATTAAGAA-TGAAGTGGCAA 1478
 Qy 458 erGlyMetIleLeuValMetAlaPheLeuIleThrTyrPheHisSerThrTyrP 478
 Db 1479 GTGGGATGATCTGGTCACTGGCTTCTTCTCTATCACCTCACCTTCACTCAACCTGGG 1538
 Qy 478 alThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGly 498
 Db 1539 TGACCACTGAGGCTACTCTTCTCCGTCACTTGTCTGTCTGCTGCTGCTGCTGCTGCTGCT 1598
 Qy 498 erArgIleIlePheAspAspPheArgGluAlaTyrTyrTyrLeuArgHisAsnThrProG 518
 Db 1599 GTAGGATCATATTTGATGACTTCCGAGAGACATATTTATTTGGCTTCTGCTATTAATCTCCAG 1658

Qy 518 luApAlaLysValMetSerTyrTyrPaspTyrGlyTyrGlnIleThrAlaMetAlaAsnA 538
 Db 1659 AGGATGCGAAGGTCATGCTCTGGTGGGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 1718
 Qy 538 rGThrIleLeuValAspAsnAsnThrTyrPaspAsnAsnThrHisIleSerArgValGlyGlnA 558
 Db 1719 GAACAATTTTAGTGGCAATACACATGATGATTAATACCCATATTTCTCGAGTAGGCAGG 1778
 Qy 558 laMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrV 578
 Db 1779 CAATGGCGTCCACAGAGAGAAAAGCCCTATGAGATCATGAGGGAGCTCGATGTGAGTATG 1838
 Qy 578 alLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuT 598
 Db 1839 TGCTGGTCATTTTGGAGGCTCTCACTGGTATTCTCTGATGATATCAACAAGTTCTTT 1898
 Qy 598 rpMetValArgIleGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrT 618
 Db 1899 GGATGGTCCGATTTGGAGGAGCACAGATACAGCAACATATCAAGGAGAAATGACTATT 1958
 Qy 618 yrThrProThrGlyLysPheArgValAspArgGluGlySerProValLeuLeuAsnCysL 638
 Db 1959 ATACTCAACTGGGAGTTCGTTGGACCGTGAAGTTCCTCAGTCTGCTCACTGCC 2018
 Qy 638 euMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProp 658
 Db 2019 TCATGTACAAGATGTGTACTATCTGCTTTGGACAGGTTTACAGAGAACCCAGCGTCC 2078
 Qy 658 rGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVal 678
 Db 2079 CAGGCTTTGACCGTTCGGAATGCTCAGATGGGAATAAAGACTTTGAGCTTGATGCTC 2138
 Qy 678 euGluGlu 680
 Db 2139 TGGAGGAA 2146

RESULT 13

HSM809254

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

CONGRITY

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 2898

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp779N2434"

/tissue_type="liver"

/clone_lib="779 (synonym: hnccl). Vector pSport1_sfi; host

DH10B; sites sfiI + SfiIb"

/dev_stage="fetal"

polyA_signal
polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 4.15e-220 Length: 2898
 Score: 220.00 Matches: 671
 Percent Similarity: 97.39% Conservative: 0
 Best Local Similarity: 97.39% Mismatches: 9
 Query Match: 31.21% Indels: 18
 DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x HSM809254 (1-2898)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 186 ATGACTAAGTTGGGATTTTTCGGATTGCTCTATGAGAGCGGACACATTTTGAAGCTT 245
 QY 21 LeuLeuSerMetAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 246 CTCATTCTGTCATGGCTGCTGTATTATCTCTCCACTCGTCTGCTGCTGCTGAGA 305
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
 DB 306 TTGAAAGTGTTATCCAGAGTTGATCCGTTTGAATTTATTCGGACTACCAAGTTCCG 365
 QY 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTrpTyrProLeu 80
 DB 366 GCTGAGGAGGGGTTTATAAATTCATAAAGTGGTTTGATGACGAGCGCTGGTACCCCTTTG 425
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 426 GGACGATCATTTGGAGGACAAATTCACAGGTTTAAATGATACCTCTCTGCTGCAATCTAC 485
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnVal-CysValPheLeuAl 120
 DB 486 CATGTACTTCCATTTTTCACATCACCATCGACATTCGGAATGC-CTGTGTGTTCTGTGC 544
 QY 120 aProLeuPheSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLys 139
 DB 545 CCCT 602
 QY 140 AspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSer 159
 DB 603 GATGACGGGGTGGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 662
 QY 160 ArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 179
 DB 663 CGATCTGGCTGGCTCTATGATATGAAGGAGTTGCCATCTTTTGCATGCTACTCACC 722
 QY 180 TyrTyrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAla 199
 DB 723 TACTACATGTGGATCAAGGAGTAAAGACATGGTTCCATCTGTGGGAGCTAAGTGTGCC 782
 QY 200 LeuAlaTyrPheTyrMetValSerTyrGlyTyrValPheLeuIleAsnLeuIle 219
 DB 783 CTTCCTTATTTACATGGTCTGCTCATGGGAGGTTATGTTTCTCTGATCACTTAAT 842
 QY 220 ProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 239
 DB 843 CCTCTCCACGTCCTCGTGGCTGATGCTCACAGGCGGCTTCTCTCACCGGATCTATGTGCC 902
 QY 240 TyrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 DB 903 TACTGTACTGTCTTACTGCTGGGCACTATATCTTCTAT-CCAGATCTCTCTTGTGGGTTT 961
 QY 259 eGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
 DB 962 CCAGCCCTGTCTCTTTCATCAGAGCACATGGCAGC-CTTTGGGGTCTTTTGTCTCTCTGCCA 1020
 QY 279 leHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnPheGluValLeuP 299

DB 1021 TCCATGCTTTTGTGGATTACCTGGCAGCAAGTTGAATCCACAACAATTTGAAGTCTTT 1080
 QY 299 heArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetL 319
 DB 1081 TCCGAGAGCGTCATCTCTGTGTAGGCTTTGTCTCTTCCACCGTGGAGCTCTCTCATGTC 1140
 QY 319 euThrGlyLysIleSerProTrpThrGlyArgPhe-TyrSerLeuLeuAspProSerTyr 338
 DB 1141 TGACAGGAAATAATCTCCCTGGACGGGCGTCT-CTACTGCTGCTGGATCCCTCTTAT 1199
 QY 339 AlaLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSer 358
 DB 1200 GCTAAGAACAACATCCCATCTCTGCTGTGTGAGCATCAGCCACACACCTGGTCC 1259
 QY 359 SerTyrTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrCys 378
 DB 1260 TCATACTATTTTGACCTGAGCTCTCTGCTCTTCATGTTTCCAGTTGGCTCTATTACTGC 1319
 QY 379 PheSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyr 398
 DB 1320 TTTAGCAACCTGCTGATGCCCGGATTTTATCATCATGTATGTTGACCGACGATGTAC 1379
 QY 399 PheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSe 418
 DB 1380 TTTTTCAGCTGTAAATGGTGGCTCTAATGCTAGTTGGACCTGTATTATGTG-CATTCTCTC 1438
 QY 418 rGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgPr 438
 DB 1439 TGGCAATTGGAGTCTCCAGTCTGCCACATACATGAAGAATCTGGACATAAGTCGTCC 1498
 QY 438 oAspLysLysSerLysGlnAspSerThrTyrProIleLysIle-GluValAlaLys 458
 DB 1499 AGCACAGAGACAGACAGACAGATTCACCTACCTCTATTAGAA-TGAAGTGGCAA 1557
 QY 458 er-GlyMetIleLeuValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrp 477
 DB 1558 GTGGGATA-ATACTGGTCATGGCTTTCTTCTTCATCACTACACCTTTCAATCAACCTGG 1616
 QY 478 ValThrSerGluAlaTyr-SerSerProSerIleValLeuSerAlaArgGlyGlyAspGly 497
 DB 1617 GTACACAGTGGAGGCTCTCTCTCTCCGTCATGTTGCTATCTGCCGCTGGTGGGAGTGGC 1676
 QY 498 SerArgIleIlePheAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrPro 517
 DB 1677 AGTAGATCATATTTGATGACTTCCGAGAGCATATATTGGCTTCGTCATAAATCTCCA 1736
 QY 518 GluAspAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsn 537
 DB 1737 GAGGATGCCAGAGTCTATGCTCTGGTGGATTTATGGCTATCAGATTACAGCTATGGCAAAC 1796
 QY 538 ArgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGln 557
 DB 1797 CGAACCAATTTTATGGAGCAATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAG 1856
 QY 558 AlaMetAlaSerThrGluLysLysAlaTyrGluIleMetArgGluLeuAspValSerTyr 577
 DB 1857 GCATGGCTCTCACAGAGGAAAGCCATGAGATCATGAGGAGCTTGTATGTCAGGTAT 1916
 QY 578 ValLeuValIlePheGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeu 597
 DB 1917 GTGCTGTCATTTTGGAGGCTCTACCTGGGTATTTCTCTGATGATATCAACAAGTTTCTT 1976
 QY 598 TrpMetValArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyr 617
 DB 1977 TGGATGGTCCGATTTGGAGGAGGACACATACAGGCAACATATCAAGGAGATGACTAT 2036
 QY 618 TyrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCys 637
 DB 2037 TATACTTCCAACTGGGAGTTCGCTGGACCGTGAAGGTTCTCCAGTGTGCTCAACTGC 2096
 QY 638 LeuMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgPro 657
 DB 2097 CTCATGTACAGATGTGTTACTATCGCTTTGGACAGGTTTACACAGAAAGCCAGCTTCT 2156

```

QY 658 ProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVal 677
Db 2157 CCAGGCTTTGACCGGTGCGAAATGCTGAGATTGGGAATAAAGACTTTGAGCTTTGATGTC 2216

QY 678 LeuGluGlu 680
Db 2217 CTGGAGGAA 2225

RESULT 14
AKI30548
LOCUS Homo sapiens cDNA FLJ27038 fis, clone SLV08213, highly similar to
DEFINITION Oligosaccharyl transferase STT3 subunit homolog.
ACCESSION AKI30548
VERSION AKI30548.1 GI:34527376
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S.,
Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Watanabe, M.,
Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S., Morinaga, M.,
Kawamura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isogai, T. and
Sugano, S.
NEDO human cDNA sequencing project
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1796)
AUTHORS Sugano, S. and Suzuki, Y.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2003) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail: fldnag@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; CDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction and 5'-end one pass sequencing: Institute of Medical
Science, University of Tokyo, Laboratory of Genome Structure, Human
Genome Center; 3'-end one pass sequencing: RAB; clone selection for
full insert sequencing: RAB and Helix Research Institute.
FEATURES
Location/Qualifiers
1..1796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SLV08213"
/tissue types="salivary gland"
/clone_lib="SLV"
/note="Cloning vector: pME18SFL3"

ORIGIN
Alignment Scores:
Pred. No.: 6.95e-191 Length: 1796
Score: 192.00 Matches: 305
Percent Similarity: 98.07% Conservative: 0
Best Local Similarity: 98.07% Mismatches: 3
Query Match: 27.23% Indels: 6
GB: 9 Gaps: 0

US-10-028-384-12 (1-705) x AKI30548 (1-1796)

QY 373 ValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleLeuTyr 392
Db 233 GTTGGCTCTATTACTGCTTTAGCAACCTGCTGATGCCCGGATTTTATCATCATGAT 292

QY 393 GlyValThrSerMetTyrPheSerAlaValMetValArgLeuMetLeuValLeuAlaPro 412
Db 293 GGTGTGACCAGCATGTACTTTTTCAGCTGTAATGGTGGCTCTAATGCTAGTGTGGCACT 352

```

```

QY 413 ValMetSer-IleLeuSer-GlyIleGlyValSerGlnValLeuSerThrTyrMetLysAs 432
Db 353 GTTATATGTG-CATTCTCTCTGGCATTTGGAGTCTCCAGGTGCTGTCCACATACATGAAGA 411

QY 432 nLeuAspIleSerArgProAspLysSerLysValSerGlnGlnAspSerThrTyrProI 452
Db 412 TCTGGACATAAGTCGTCACAGCAAGAGAGCAAGAGCAAGAGGATTCCACCTACCCCTAT 471

QY 452 elysIle-GluValAlaSerGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrT 472
Db 472 TAAGAA-TGAAGTGGCAAGTGGCATGATACTGGTCATGGCTTCTTTCTCATCACCTACA 530

QY 472 hrPheHisSerThrTyrValThrSerGluAlaTyrSerSerProSerIleValLeuSerA 492
Db 531 CTTTTCATTCAACCTCGGTGGACGAGGAGGCTCTCTCTCGTCCATTGACTATCTG 590

QY 492 laArgGlyGlyAspGlySerArgIleIlePheAspAspPheArgGluAlaTyrTyrL 512
Db 591 CCGGTGGTGGGATGGCATAGGATCATATTGATGACTTCCGAGACGATATTATTATGGC 650

QY 512 euArgHisAsnThrProGluAspAlaLysValMetSerTyrTrpAspTyrGlyTyrGlnI 532
Db 651 TTTCGTCATAATACTCCAGAGGATGCGAAGGTCATGCTCTGTGGGATTATGGCTATCAGA 710

QY 532 leThrAlaMetAlaAsnArgThrIleLeuValAspAsnAsnThrTyrAsnAsnThrHisI 552
Db 711 TTACGACTATGGCAACCGACCAATTTAGTGGACATAACACATGGGAATTAATACCCATA 770

QY 552 leSerArgValGlyGlnAlaMetAlaSerThrGluGluLysAlaTyrGluIleMetArgG 572
Db 771 TTTCGAGTAGGGGAGGCAATGGCGTCCACAGAGAGAAAAAGCCCTATCAGATCATGAGG 830

QY 572 luLeuAspValSerTyrValLeuValIlePheGlyGlyLeuThrGlyTyrSerSerAspA 592
Db 831 AGCTCGATGTGAGCTATGCTGTGTCATTTTGGAGGCTCAGTGGGATTTCCTCTCTGATG 890

QY 592 spIleAsnLysPheLeuTyrMetValArgIleGlyGlySerThrAspThrGlyLysHisI 612
Db 891 ATATCAACAAGTTCTTTGGATGGTCCGGATTGGAGGAGCACAGATACAGCGCAACATA 950

QY 612 leLysGluAsnAspTyrTyrThrProThrGlyGluPheArgValAspArgGluGlySerP 632
Db 951 TCAGAGCAATGACTATTATATCTCAACTGGGGAGTTCCCGTGGACCGTGAAGTTCTC 1010

QY 632 roValLeuLeuAsnCysLeuMetTyrLysMetCysTyrTyrArgPhe-GlyGlnValTyr 651
Db 1011 CAGTCTGCTCAACTGCTCATGTACAAAGATGTGTACTATCGCTG-TGGACAGGTTTAC 1059

QY 652 ThrGluAlaLysArgProProGlyPheAspArgValArgAsnAlaGluIleGlyAsnLys 671
Db 1070 ACAGAGCAAGCGTCTCTCCAGGCTTTGACCGTGTCCGAAATGTGTGAGATTGGGAATAA 1129

QY 672 AspPheGluLeuAspValLeuGluGlu 680
Db 1130 GACTTTGAGCTTTGATGCTCTGGAGGAA 1156

```

```

RESULT 15
HUNTMTC Homo sapiens transmembrane protein (TMC) mRNA linear PRI 13-FEB-2002
LOCUS L47337
DEFINITION Homo sapiens transmembrane protein (TMC) mRNA, complete cds.
ACCESSION L47337
VERSION L47337.1 GI:18654193
KEYWORDS transmembrane protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2293)
Bellacosa, A.A.B. and Testa, J.J.R.
REFERENCE Isolation, characterization and mapping to human chromosome
AUTHORS 1q24-25 of a gene, TMC, encoding a highly conserved putative
TITLE transmembrane protein

```


JOURNAL
FEATURESUnpublished
Location/Qualifiers

1..2293
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="11924-25"
/cell_type="mesothelioma"
/clone_lib="ddcll"
1..2293
/gene="TMC"
117..2234
/gene="TMC"
/note="putative"
/codon_start=1
/product="transmembrane protein"
/protein_id="AAU77539.1"
/db_xref="GI:18654134"
/translation="MTKFGFRLSYEKQDTLLKLILSMAVLSFSTRFAVLRFSV
IHEEDPYNRTFRFSEGFYKFNWFDRAWYPLGRIIGGTIYFGLMITSAAIYHV
LHPHITIDIRNVCVFLAPLFSFTTIVYHLTKELKADAGLLAAALIAVVPYISR
SVAGSYNEGIAIPLMLTYMMIKAVKTGSIQWAAKACALAYFYVSVWGGVFLINL
IPLHVLVLMRLSGFSPHRIYVAYCTVYCLGTILSMQISFVGPQVPLSSHMAAFVGL
COIHAVDYLRSLNPOQPEVLRSVTSVGLVLLTVGALLMTGKISPTWGRFVSL
DPSYAKNIPILASVSEHPTWSSVYFQLQVLFVFPVGLYCFNSISDARIPLIMY
GYTSYSAWVRLMLVLRVCLISGLIGVSQVLSYMNLOISAPDKSKKQDDSTY
PIKEVASGMILVNAFLITITTHSTWTSSEYSSPISLARSDDGSGRIIPDDPREA
YYUHNHTPDAKVMWMDYGYQITAMANRTILVDNNTWNTTHISRGQANASTBEKA
YEIMREDVSLVILFGLGTGYSDDINKFLMVRIGGSTDTGKHIRENDYVTPGEF
RVDRGSPVLLNCLMYKCYRFGQVYTEAKRPPGPDVRNAEIGNKDFELDLVEAY
TTEHVLVRIYKVKLDNRGLSRT"

ORIGIN

Alignment Scores:

Pred. No.: 1 72e-184 Length: 2293
Score: 186.00 Matches: 672
Percent Similarity: 97.67% Conservative: 0
Best Local Similarity: 97.67% Mismatches: 8
Query Match: 26.38% Indels: 16
DB: 9 Gaps: 0

US-10-028-384-12 (1-705) x HUMTMC (1-2293)

QY 1 MetThrIysPheGlyPheLeuArgLeuSerTyGluLysGlnAspThrLeuLeuLysLeu 20
DB 117 ATGACTAAGTTTGGATTTCGCGATTGCTCTATGAGAACGAGACACATTTTGAAGCTT 176
QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
DB 177 CTCATTCTGTCATGCGTCTGTATATCTTCTCCACTCGTCTGTTGCTGCTGAGA 236
QY 41 PheGluSerValIleHisGluPheAspProTyPheAsnTyArgThrThrArgPheLeu 60
DB 237 TTTGAAAGTGTATCCATGATGTTGATCCGCTACTTAAATATCGGACTACCAAGTTCCTG 296
QY 61 Ala-GluGluGlyPheTyLysPheHisAsnTTPheAspAspArgAlaTrpTyProle 80
DB 297 TC-TGAGAGAGGGTATTAATAATCCATACTCGTTTGTATGACCGAGCTGTACCCCTT 355
QY 80 uGlyArgIlelleGlyGlyThrIleTyProGlyLeuMetIleThrSerAlaAlaIleTy 100
DB 356 GGGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATCACCCTCTGCTGCAATCTA 415
QY 100 rHisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAl 120
DB 416 CCATGACTCATTTTTCACATCACCATCGACATTCGGAATGCTGTGTGTTCTGCGC 475
QY 120 aProLeuPheSerSerPheThrSer-IleValThrTyLysLeu-LeuThrLysGluLeuLys 139
DB 476 CCCT 533
QY 140 AspAlaGlyAlaGlyLeuLeuAlaAlaAlaMetIleAlaValValProGlyTyLysSer 159

DB 534 GATGCAAGGGCTGGGCTTCTTCTGCTGCCATGATTGCTGTAGTTCCTGGATATATCTCC 593
QY 160 ArgSerValAlaGlySerTyArgPheGlnGluGlyIleAlaIlePheCysMetLeuLeuThr 179
DB 594 CGATCTGTGGCTGGCTCTATGATATGAAGGGATTGCCATCTTTTGCATGCTACTCACC 653
QY 180 TyTyMetTrpIleLysAlaValLysThrGlySerIleCysTTPAlaAlaLysCysAla 199
DB 654 TACTACATGTGGATCAAGGCACGTAAAGACTGGTTCATCTGTTGGCAGCTAAGTGTGCC 713
QY 200 LeuAlaTyPheTyMetValSerSerTrpGlyGlyTyValPheLeuLeuLeuLeu 219
DB 714 CTTGCTTATTTCTACATGGTCTCGTCATGGGAGGTTATGTTCTTCATCACTTAAT 773
QY 220 ProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyValAla 239
DB 774 CTTCTCCAGTCTCTGCTGCTGATGCTCACAGGCCGTTTCTCTCACCGGATCTATGTGCC 833
QY 240 TyCysThrValTyCysLeuGlyThrIleLeuSerArg-GlnLeuSerPheValGlyPhe 259
DB 834 TACTGTACTGTTTACTGCTGGGCACTATCTTTCTAT-GCAGATCTCTCTTTGGGTTT 892
QY 259 eGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
DB 893 CCAAGCTCTCTTCTTATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAGA 951
QY 279 IeHisAlaPheValLeuSerTyLeuArgSerLysLeuAsnProGlnGlnPheGluValLeu 299
DB 952 TCCATGCTTTTGGATTACCTGCGCAGCAAGATTGAATCCCAACAATTTGAAGTCTTT 1011
QY 299 heArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyValAlaLeuMetL 319
DB 1012 TCCGAGAGCTCATCTCTGCTGAGCTTTTGTCTTCTCACCTGGAGCTCTCTCTATGC 1071
QY 319 eutThrGlyLysLeuSerProThrThrGlyArgPheTySerLeuLeuAspProSerTyArg 339
DB 1072 TGACAGAGAAAATATCTCTCCCTGGACGGGGCGTTCCTACTCGCTCTGGATCCCTCTTATG 1131
QY 339 lalysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerS 359
DB 1132 CTAGAGAACACATCCCATCATGCTCTGCTGTGTGTGAGCATGCCACACACCTGGTCT 1191
QY 359 erTyTyThrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyTyCysP 379
DB 1192 CATACTATTTTACCTGCGAGCTCTCTGCTCTCATGTTTCCAGTTTGGGCTCTTATCTGCT 1251
QY 379 heSerAsnLeuSerAspAlaArgIlePheIleIleMetTyGlyValThrSerMetTyP 399
DB 1252 TTAGCAACCTGTCTGATGCCGGATTTTATCATCATGTATGTTGTGACGAGCATGTACT 1311
QY 399 heSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSer 418
DB 1312 TTTCACTGTAAATGTTGGTCTTAATGCTAGTGTGGACCTGTATGTG-CATTCTCTCT 1370
QY 419 GlyIleGlyValSerGlnValLeuSerThrTyTyMetLysAsnLeuAspIleSerArgPro 438
DB 1371 GGCATTGGAGTCTCCCGAGTGTCTGCACATACATGAAGATCTGGACATAGTCGCCCA 1430
QY 439 AspLysLysSerLysGlnGlnAspSerThrTyProIleLysIle-GluValAlaIle 458
DB 1431 GACAAGAGAGCAAGAACACAGGATTTCCACTACCTACCTATTAAGAA-TGAAGTGGCAAG 1489
QY 458 rGlyMetIleLeuValMetAlaPhePheLeuIleThrTyThrPheHisSerThrTrpVa 478
DB 1490 TGGATGATGACTGTGCTGCTTCTTCTCATCACCTACACCTTTTCATTCACTGGGT 1549
QY 478 lThrSerGluAlaTySerSerProSerIleValLeuSerAlaArgGly-GlyAspGlyS 498
DB 1550 GACCAGTAGGGCTCTCTTCTCCGTCATGTTACTTCTGCCCGTGA-TGGGGATGGCA 1608
QY 498 exArgIleIlePheAspAspPheArgGluAlaTyTyThrIleArgHisAsnThrProG 518
DB 1609 GTAGGATCATATTTGATGACTTCCGAGAGCATATTTATGCTTCGTCATAATACTCCAG 1668

```
Qy 518 luAspAlaLysValMetSerTirpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnA 538
Db 1669 AGGATGCCAAGGTCAATGCTGGTGGATTATGGCTATCAGATTACAGCTATGGCAACC 1728
Qy 538 rgThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnA 558
Db 1729 GAACAAATTTTGTGGACAAATAACACATGGAAATAATACCCATATTTCTCGAGTAGGGCAGG 1788
Qy 558 laMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrV 578
Db 1789 CAATGGCGCTCCACAGAGGAAAGCCCTATGAGATCATGAGGAGCTCGATGTGAGCTATG 1848
Qy 578 alleuValIlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeut 598
Db 1849 TGCTGGTCATTTTGGAGGCCCTCACTGGGTATTCCTCTGATGATATCAACAAGTTTCCTT 1908
Qy 598 rpMetValArgIleGlyGlySerThrAspThrGlyLysHisIleLysGluAsnAspTyrT 618
Db 1909 GGATGGTCCGGATTGGAGGGAGCACACATACAGGCAACATATCAAGGAGAAATGACTATT 1968
Qy 618 yrThrProThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysL 638
Db 1969 ATACTCCAACCTGGGAGTTCGGTGTGGACCGTGAAGGTTCCTCAGTGTCTCAACTGCC 2028
Qy 638 euMetTyrLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProp 658
Db 2029 TCATGTACAGATGTGTACTATCGCTTTGGACAGGTTTACAGAGGCCAAGCGTCCTC 2088
Qy 658 roGlyPheAspArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspVall 678
Db 2089 CAGGCTTTGACCGTGTCCGAAATGCTGAGATTGGGAATAAGACTTTGAGCTTGATGTCC 2148
Qy 678 euGluGlu 680
Db 2149 TGGAGGAA 2156
```

Search completed: December 14, 2004, 22:48:51
Job time : 7872.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 812.278 Seconds
(without alignments)
4556.130 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 MTKFGFLSLSEKQDTLLKL:.....HMLVRIYKVDLDRGLSRT 705

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 1

Total number of hits satisfying chosen parameters: 8266295

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+pin.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10028384/runat_14122004_131645_6790/app_query.fasta_1.3740
-DB=N_Geneseq_23Sep04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384 @CGN 1.1.1986 @runat_14122004_131645_6790 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 705 | 100.0 | 2472 | 10 | ADD94793 |
| 2 | 705 | 100.0 | 2472 | 10 | ADH28844 |
| 3 | 226 | 32.1 | 2760 | 5 | ABV24502 |
| 4 | 152 | 21.6 | 3094 | 10 | ADD94791 |
| 5 | 146 | 20.7 | 440 | 6 | ABK27622 |
| 6 | 140 | 19.9 | 455 | 6 | ABL37867 |

ALIGNMENTS

RESULT 1
ADD94793
ID ADD94793 standard; DNA; 2472 BP.

XX

AC ADD94793;

XX 29-JAN-2004 (first entry)

DE Human ITM1 gene sequence.

XX source of immunodominant MHC-associated peptide; SIMP; MHC;

XX major histocompatibility complex; human leukocyte antigen; HLA;

XX cytostatic; immunosuppressive; antineoplastic; gene therapy; cancer;

XX lung cancer; intestine cancer; sarcoma; prostate cancer;

XX testicular cancer; breast cancer; melanoma; pancreatic cancer;

XX haematological cancer; immune response; lymphoid cell proliferation;

XX autoimmune disease; transplant rejection; SIMP-derived peptide; human;

XX gene; ds; ITM1.

XX Homo sapiens.

XX WO2003054008-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-CA001967.

XX 20-DEC-2001; 2001US-00028384.

XX (COMP-) COMPATIGENE INC.

Aa16208 Human col
Adp28508 Human sec
Aac98825 Human pan
Aac00273 Human sec
Abk39617 cDNA enco
Aca11946 Human lun
Aca03132 Lung carc
Adh47174 Human lun
Aaf68440 Human lun
Abk39233 DNA enco
Abk38351 cDNA enco
Aca11562 Human lun
Aca10680 Human lun
Abx99631 Lung carc
Aca02748 Lung carc
Adh46790 Human lun
Adh45894 Human lun
Ade1646 Human lun
Abx45128 Bovine ES
Abn73253 Bovine em
Abn73343 Bovine em
Ach34952 Human end
Aaf68653 Human lun
Abk38564 cDNA enco
Aca10893 Human lun
Abx9844 Human carc
Adh46107 Human lun
Ade1859 Human lun
Abq55493 Human ova
Aah83346 Human ova
Aas24717 Human ova
Abx35963 Bovine ES
Abx41175 Bovine ES
Abq58556 Human col
Abi38310 Human col
Abq59387 Human col
Abi02795 Drosophil
Abi102794 Drosophil
Abv86822 Human col

7 128 18.2 487 3 AAA16208
8 127 18.0 1371 12 ADP28508
9 126 17.9 575 3 AAC98825
10 102 14.5 414 3 AAC00273
11 102 14.5 487 6 ABK39617
12 102 14.5 487 8 ACA11946
13 102 14.5 487 8 ACA03132
14 102 14.5 487 10 ADH47174
15 101 14.3 307 5 AAF68440
16 101 14.3 307 6 ABK39233
17 101 14.3 307 6 ABK38351
18 101 14.3 307 8 ACA11562
19 101 14.3 307 8 ACA10680
20 101 14.3 307 8 ABX99631
21 101 14.3 307 8 ACA02748
22 101 14.3 307 10 ADH46790
23 101 14.3 307 10 ADH45894
24 101 14.3 307 12 ADE1646
25 94 13.3 419 8 ABX45128
26 81 11.5 596 6 ABN73253
27 81 11.5 596 6 ABN73343
28 80 11.3 476 9 ACH34952
29 65 9.2 307 5 AAF68653
30 65 9.2 307 6 ABK38564
31 65 9.2 307 8 ACA10893
32 65 9.2 307 8 ABX9844
33 65 9.2 307 10 ADH46107
34 65 9.2 307 12 ADE1859
35 59 8.4 573 6 ABQ55493
36 54 7.7 526 5 AAH83346
37 53 7.5 516 4 AAS24717
38 42 6.0 261 8 ABX35963
39 42 5.0 382 8 ABX41175
40 40 5.7 473 6 ABQ58556
41 40 5.7 515 6 ABL38310
42 40 5.7 665 6 ABQ59387
43 39 5.5 2855 4 ABL02795
44 39 5.5 6153 4 ABL02794
45 38 5.4 243 6 ABV86822

| | | | |
|----------|---|---|------|
| Db | 1607 | ATATTGTGAGCTCTCCGAGAAGCATATATTGCGTTCGTCATPAATACTCCAGAGATGCG | 1666 |
| Qy | 521 | LysValMetSerTrrTpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIle | 540 |
| Db | 1667 | AAGGTTCATGCTCGTGTGGGATTATGGGTATCAGATTACAGCTATGCGAAACCGAACATT | 1726 |
| Qy | 541 | LeuValAspAsnAsnThrTrrPAsnAsnThrHisIleSerArgValGlyGlnAlaMetAla | 560 |
| Db | 1727 | TTAGTGGACAATAACACATGGAAATAATACCCATATTCTCGAGTAGGGCAGGCATGGCG | 1786 |
| Qy | 561 | SerThrGluGluLysAlaTyrGluIleMetArgGluLeuLeuAspValSerTyrValLeuVal | 580 |
| Db | 1787 | TCCACAGAGGAAAGCCATATGAGATCATGAGGAGCTCGATGTCAAGCTATGTGCTGTC | 1846 |
| Qy | 581 | IlePheGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrrMetVal | 600 |
| Db | 1847 | ATTTTGGAGGCTCACCTGGGTATTCTCTGATGATATCAACAAGTTCTTTGGATGGTC | 1906 |
| Qy | 601 | ArgIleGlyGlySerThrAspThrGlyValHisIleLysGluAsnAspTyrTrrThrPro | 620 |
| Db | 1907 | CGGATTGGAGGAGCAGACATACAGGCAACATATCAAGGAGAATGACTATTATACTCCA | 1966 |
| Qy | 621 | ThrGlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyr | 640 |
| Db | 1967 | ACTGGGGAGTTCCGTGTGGACCGTGAAGGTTCTCCAGTGTCTCACTGCTCATGTAC | 2026 |
| Qy | 641 | LysMetCysTyrTrrArgPheGlyGlnValTyrThrGluAlaLysArgProProGlyPhe | 660 |
| Db | 2027 | AGATGTGTACTATCGCTTTGGACAGGTTTACACAGAAGCCAAAGCGTCTCCAGGGCTT | 2086 |
| Qy | 661 | AspAsgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu | 680 |
| Db | 2087 | GACCGTCCGAAATGCTGACATTGGCAATAAAGACTTTTGAGCTTGATGCTCTGGAGAA | 2146 |
| Qy | 681 | GlyTyrThrThrGluHisTrrPLeuValArgIleTyrLysValLysAspLeuAspAsnArg | 700 |
| Db | 2147 | GGCTATACACAGAACATTGGCTGGTCAGGATATACAAGTAAGAGACCTGGATAATCGA | 2206 |
| Qy | 701 | GlyLeuSerArgThr 705 | |
| Db | 2207 | CGCTGTCAAGGACA 2221 | |
| RESULT 2 | | | |
| ADH28844 | | | |
| ID | ADH28844 standard; DNA; 2472 BP. | | |
| XX | ADH28844; | | |
| AC | XX | | |
| XX | 11-MAR-2004 (first entry) | | |
| DT | XX | | |
| DE | Human chronic myclogenous leukaemia (CML) gene marker #112. | | |
| DE | XX | | |
| XX | de; chronic phase chronic myelogenous leukaemia; CP-CML; | | |
| XX | blast crisis CML; BC-CML; human; chronic myclogenous leukaemia; | | |
| KW | gene marker. | | |
| OS | Homo sapiens. | | |
| XX | US200310426-A1. | | |
| PN | XX | | |
| PD | 05-JUN-2003. | | |
| XX | 14-JUN-2002; 2002US-00171591. | | |
| PF | XX | | |
| PR | 18-JUN-2001; 2001US-0298914P. | | |
| XX | (LINSLEY P S. | | |
| PA | (MAOM/) MAO M. | | |
| PA | (DAIH/) DAI H. | | |
| PA | (HEYI/) HE Y. | | |
| PA | (RADJ/) RADICH J P. | | |
| XX | Linsley PS, Mao M, Dai H, He Y, Radich JP. | | |

CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX

SQ Sequence 2760 BP; 688 A; 645 C; 658 G; 769 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4, 1e-224 Length: 2760
 Score: 226.00 Matches: 674
 Percent Similarity: 97.9% Conservative: 0
 Best Local Similarity: 97.9% Mismatches: 6
 Query Match: 32.06% Indels: 14
 DB: 5 Gaps: 0

US-10-028-384-12 (1-705) x ABV24502 (1-2760)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 105 ARGACTAAGTTGGGATTTTGGGATTTGCGATTGTCCTATGAGAGCAGGACACATTTTGAAGCTT 164
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeu-PheAlaValLeuAr 40
 DB 165 CTCATCTGTCATAGGCTGCTGATTATCCCTTCCTCCACTGCTGTTTGCTGTCCTGAG 224
 QY 40 gPheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLe 60
 DB 225 ATTTGAAAGTGTATCCATGAGTTGATCCGCTACTTTAAATATCCGACTACCGATTCTCT 284
 QY 60 uAlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspArgAlaTyrProLe 80
 DB 285 GGCTGAGAGGGGTTTTATAATTTCCATTAACCTGTTTGATGACCGCTGGTACCCCTTT 344
 QY 80 uGlyArgIleleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTy 100
 DB 345 GGGACGAATCATGAGAGACAAATTTACCCAGGTTTAAATGATCACTCTGTCGCAATCTA 404
 QY 100 rHisValLeuHis-PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuA 120
 DB 405 CCATGACTCCATTTTTTCCACATCCACATCGACATCGGAATGCTGTGTGTCTCTGG 464
 QY 120 laProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLy 139
 DB 465 CCCCTCTCTCTCTCTCTTACCAC-CATGTCACGTACCA-CCTTACCAAGAGGTCAA 522
 QY 139 sAspAlaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValProGlyTyrIleSe 159
 DB 523 GGATGCAAGGGGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 582
 QY 159 rArgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuTh 179
 DB 583 CCGATCTGTGGTGGCTCCTATGATGATGAGGGATGGCATCTTTTGGCATGCTACTCAC 642
 QY 179 rTyTyrMetThrIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAl 199
 DB 643 CTACTACATGTGGATCAAGCAGTAAAGACTGGTTCCTATCTGTTGGCAGTAAAGTGC 702
 QY 199 aLeuAlaTyrPheTyrMetValSerSerTrpGlyTyrValPheIleuLeuLeuI 219
 DB 703 CCTTGTCTTATTTTACATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
 QY 219 eProLeuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAl 239
 DB 763 TCCTCTCCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 822
 QY 239 aTyrCysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyP 259
 DB 823 CTACTGACTGTTTACTGCTGGGCACTACTTTCTAT-GCAGATCTCTTTTGGGTT 881
 QY 259 heGlnProValLeuSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGln 278
 DB 882 TCCAGCTGCTCTTTCATCAGACACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAG 940

QY 279 IleHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeu 298
 DB 941 ATCCATGCCCTTGTGGATTACCTGCGCAGCAAGTTGAATCCACACAATTTGAAGTTCTT 1000
 QY 299 PheArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMet 318
 DB 1001 TTCGGAGCGTCACTCTCTGGTAGGCTTGTCTCTTCCCGTGGAGCTCTCTCTCATG 1060
 QY 319 LeuThrGlyLysIleSerProTrpThrGlyArgPheTyr-SerLeuLeuAspProSerTyr 338
 DB 1061 CTGACAGGAAAATAATCTCTCGACGGGGGCTTCTACTCGTGTGTGATCCCTCTTAT 1120
 QY 339 AlaLysAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSer 358
 DB 1121 GCTAAGAACAAACATCCCATCTGCTTCTGTGTCTGAGCATCAGCCCAACCTGGTCC 1180
 QY 359 SerTyrTyr-PheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyrCys 378
 DB 1181 TCATCTATTTTGACCTGCGCTCTCTGCTCTTCACTGTTTCCAGTTGGCCTCTATTACTGC 1240
 QY 379 PheSerAsnLeuSerAspAlaIlePheIleIleMetTyrGlyValThrSerMetTyr 398
 DB 1241 TTTAGCAACCTGCTGATCCCGGATTTTATCATCATCTGATGTTGTGACCAAGCATGTAC 1300
 QY 399 PheSerAlaValMetValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSe 418
 DB 1301 TTTTCAGCTGTAATGGTGGCTTAATGCTAGTGTGGACCTGTTATGTG-CATTCTCTTC 1359
 QY 418 rGlyIleGlyValSerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgPr 438
 DB 1360 TGCGATTTGAGTCTCCAGGTCTGTCCACATACATGAAGAATCTGGACATTAAGTCGTC 1419
 QY 438 oAspLysLysSerLysGlnGlnAspSerThrTyrProIleLysIle-GluValAlaAs 458
 DB 1420 AGACAGAAGAGCAAGAACAGATTCACCTACCTATACCAATTAAGAA-TGAAGTGGCAA 1478
 QY 458 exGlyMetIleLeuValMetAlaPhePheLeuIleThrTyrThrPheHisSerThrTrpV 478
 DB 1479 GTGGGATGATATGCTGCTGCTGCTTCTTCTCATCATCACCTACACCTTTTCATCAACCTGG 1538
 QY 478 alThrSerGluAlaTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlys 498
 DB 1539 TGACAGTGGAGCCCTACTCTTCTCCGTCCTATGTTCTGTCTGCCGCTGTGGGATGGCA 1598
 QY 498 exArgIleIlePheAspAspPheArgGluAlaTyrTyrTyrPheHisSerThrTrpV 518
 DB 1599 GTAGGATCATATTTGATGACTTCCGAGAGCATATTAATTTGGCTTCGTCAATTAATCTCCAG 1658
 QY 518 luAsAlaLysValMetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnA 538
 DB 1659 AGGATGCCAAGGTCATGTCCTGGTGGGATTTATGGCTATCAGATTACAGCTATGGCAACC 1718
 QY 538 rGThrIleLeuValAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnA 558
 DB 1719 GBAACAATTTTAGTGACAAATAACATGGAATAATACCATATTTCTCGAGTAGGGCAGG 1778
 QY 558 laMetAlaSerThrGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrV 578
 DB 1779 CAATGGCTCCACAGAGAGAAAAGCTATGAGATCATGAGGAGCTGATGTCAGCTATG 1838
 QY 578 alLeuValIlePheGlyLysLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuT 598
 DB 1839 TGCTGGTCATTTTGGAGGCTCTACTGGTATTCCTCTGATGATATCAACAAGTTTCTTT 1898
 QY 598 rpMetValArgIleGlyLysSerThrAspThrGlyLysHisIleLysGluAsnAspTyrT 618
 DB 1899 GGATGCTCGGATTTGGAGGAGCAGATACAGGCAACATATCAAGAGAGNAATGACTATT 1958
 QY 618 yrThrProThrGlyLysPheArgValAspArgGluGlySerProValLeuLeuAsnCysL 638
 DB 1959 ATATCCAACTGGGAGTTCCGTGTGGACCGTGAAGGTTCTCCAGTCTCTCAACTGCC 2018

QY 638 euMetTyLysMetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProp 658
 Db 2019 TCATGTACAAGATGTTACTATCGCTTTGGACAGGTTTACACAGAGGCGAGGCTCTC 2078
 QY 658 rGlyPheAspArgValArgAsnAlaGluLeuGlyAsnLysAspPheGluLeuAspVal 678
 Db 2079 CAGGCTTTGACCGGTGTCGGAATCTGAGATTGGGAATAAGACTTTTGAGCTTGATGTC 2138
 QY 678 euGluGlu 680
 Db 2139 TGGAGGA 2146

RESULT 4
 ADD94791
 ID ADD94791 standard; DNA; 3094 BP.
 XX
 AC ADD94791;
 DT 29-JAN-2004 (first entry)
 XX
 DE Mouse ITM1 gene sequence.
 XX
 KW source of immunodominant MHC-associated peptide; SIMP; MHC;
 KW major histocompatibility complex; human leukocyte antigen; HLA;
 KW cytostatic; immunosuppressive; antisense therapy; gene therapy; cancer;
 KW lung cancer; intestine cancer; sarcoma; prostate cancer;
 KW testicular cancer; breast cancer; melanomas; pancreatic cancer;
 KW haematological cancer; immune response; lymphoid cell proliferation;
 KW autoimmune disease; transplant rejection; SIMP-derived peptide; mouse;
 KW murine; Gene; ds; ITM1.
 XX
 OS Mus musculus.
 XX
 PN WO2003054008-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 18-DEC-2002; 2002WO-CA001967.
 XX
 PR 20-DEC-2001; 2001US-00028384.
 XX
 PA (COMP-) COMPATIGENE INC.
 PI Perreault C, McBride K;
 XX
 DR WPI; 2003-559122/52.
 DR P-PSDB; ADD94792.
 XX

New human source of immunodominant MHC-associated peptide (SIMP) nucleic acids and proteins, useful for diagnosing and treating cancers, e.g. lung or breast cancer, or for suppressing an immune response in an autoimmune disease.

Disclosure; SEQ ID NO 9; 66pp; English.

This invention relates to a novel isolated or purified human protein, termed source of immunodominant major histocompatibility complex (MHC) - associated peptide (SIMP), which is expressed ubiquitously in human cells, where the protein has the potential of generating several protein fragments binding with high affinity to a human leukocyte antigen (HLA) molecule. The invention may allow development of therapeutics with cytostatic or immunosuppressive activity or provide sequences useful for antisense therapy or gene therapy. The source of immunodominant MHC-associated peptide (SIMP) nucleic acids, proteins and fragments are useful for diagnosing and treating cancers, for example lung cancer, intestine cancer, sarcomas, prostate cancer, testicular cancer, breast cancer, melanomas, pancreatic cancer or haematological cancer. The SIMP proteins are also useful for modulating an immune response. Decreasing lymphoid cell proliferation is useful for suppressing an immune response responsible for an autoimmune disease or a transplant rejection. The present sequence is that of the mouse ITM1 gene which is related to the invention. Note: This sequence does not appear in the specification but was obtained by the indexer from GenBank.

XX
 SQ Sequence 3094 BP; 758 A; 651 C; 703 G; 982 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,786-147 Length: 3094
 Score: 152.00 Matches: 667
 Percent Similarity: 97.37% Conservative: 0
 Best Local Similarity: 97.37% Mismatches: 9
 Query Match: 21.56% Indels: 18
 DB: 10 Gaps: 0
 US-10-028-384-12 (1-705) x ADD94791 (1-3094)
 QY 5 GlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeuLeuLysLeuSer 24
 Db 124 GGATTTTGGATTGCTTATGAGAGGAGGACACACTTCTAAAGCTTCTCATCTGCTCG 183
 QY 25 MetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArgPheGluSerVal 44
 Db 184 ATGGCTGCTGTGTTATCTTTTCTCTCTTTTGTGCTGAGATTGAAAGTCT 243
 QY 45 IleHisGluPheAspProTyrPheAsnTyrArgThrThrA:gpheLeuAlaGluGly 64
 Db 244 ATCCATGAGTTTGCATCCGTACTTTAATATCGGACTACCCGGTTTCTGGCTGAGGAGGG 303
 QY 65 PheTyrLysPheHisAsnTrpPheAspAspArgAlaTyrTyrProLeuGlyArgLeu 84
 Db 304 TTTTATAATTCCATTAACGGTTTGTATGACCGGGCTTGTACCTTTGGGCGGATCAT 363
 QY 85 GlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyrHisValLeuHis 104
 Db 364 GGAGGAACAATTTACCCAGGTTTAAATGATCACTTCTGCTGCAATCTACCATGTACTCCAT 423
 QY 105 PhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAlaProLeuPheSer 124
 Db 424 TTTTTCATATCATTATGACATTCGGAATGCTGTGTTTCTTGGGCCCTCTTCTTCTCC 483
 QY 125 SerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysAspAlaGlyAla 144
 Db 484 TCTTTTCAACAC-CATCGTTACGTATCA-CCTTTACCAAGAGAGCTCAAGGATCAGAGAGCTG 541
 QY 144 LysLeuAlaAlaAlaMetIleAlaValValProGlyTyrLysSerArgSerValAla 164
 Db 542 GGCTTTCTGCTGCTGCCATGATTTGCTGATTTCTTGGGTATATTTCTGCATCTGAGCTG 601
 QY 164 LysTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThrTyrTyrMetTrp 184
 Db 602 GCTCTCATATAATGAAGGAATTTGCTATCTTTTGCATGCTGCTTACTTACTTACTGTGGA 661
 QY 184 LysAlaValLysThrGlySerIleCys-TpAlaAlaLysCysAlaLeuAlaTyr-Phe 203
 Db 662 TCAAGGAGTGAAGACTGCTTCCATCTA-TTGGGCTGCCAAGTGTGCTCTGCTTATTTTC 720
 QY 204 TyrMetValSerSerTrpGlyGlyTyrValPheLeuLeuAsnLeuLeuProLeuHisVal 223
 Db 721 TACATGGTCTCTTCATGGGAGGCTATGTTTCTGTATCAACTTGATTCCTCTTACTATGTC 780
 QY 224 LeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAlaTyrCysThrVal 243
 Db 781 CTGGTCTTAATGCTGACAGGCGGTTTTCTTCTACCGGATCTAGTAGCTTACTTACTGTT 840
 QY 244 TyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPheGlnProVal 263
 Db 841 TACTGCTGGGACCACTTCTTCTAT-GCAGATTTCCTTTGTGTGTTTTCAGCCCGCTCT 899
 QY 263 uSerSerGluHisMetAlaGly-PheGlyValPheGlyLeuCysGlnIleHisAlaPheV 283
 Db 900 TTTATCAGAACACATGCGAGC-CTTTGGAGTGTTTGTCTCTGTGAGATCCATGCTTTCG 958
 QY 283 AlaPyrLeuArgSerLysLeuAsnProGlnPheGluValLeuPheArgSerVal 303
 Db 959 TAGATTACCTGGGAGCAAGTTGAATCCACAGCAATTCGAAGTCTCTTTTCCGGAGTGT 1018

QY 303 leSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuLeuMetLeuThrGlyLysI 323
Db 1019 TCTCCCTGGTTGGCTTTGCTCCTCCTCACTGTGGGAGCTCTCCATCATCAAGGAA 1078
QY 323 leSerProThrThrGlyArgPheThrSerLeuLeuAspProSerTyrAlaLysAsnI 343
Db 1079 TTCTCCCTGGACAGGGGCTTTCTACCTCTGCTGGATCCCTCTTATGCTAAGAA 1138
QY 343 leProIleAlaSerValSerGluHisGlnProThrThrTrpSerSerTyrPheA 363
Db 1139 TTCCCATATTGCACTCTTTCTGAGCAGCAGCCACACCTGCTCTTCTACTATT 1198
QY 363 spLeuGlnLeuLeuValPheThrPheProValGlyLeuTyrTyrCysPheSerAsnLeu 383
Db 1199 ATCTACAGCTCTTCTGCTTCTATGTTTCCAGTTGGCTCTTATCTGCTTTAGCA 1258
QY 383 exAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPheSerAlaVal 403
Db 1259 CTGATGCTCGGATTTTATCATCATGATGATGCTGACCATGATCTTTTCAGCTGA 1318
QY 403 etValArgLeuMetLeuValLeuAlaProValMetSer-IleLeuSerGlyIleGlyVal 422
Db 1319 TGGTGGCTTAATGCTGATGATGGCAGCTGTTATGTC-CATCTCTTCTGGATGGGT 1377
QY 423 SerGlnValLeuSerThrTyrMetLysAsnLeuAspIleSerArgProAspLysLysSer 442
Db 1378 TCCAGGTCCTGCTCACATATATGAANAATCTGGACATAGTCGCCCAGACAAAGAGC 1437
QY 443 LysLysGlnAspSerThrTyrProIleIleIleGluValAlaSerGlyMetIleIle 462
Db 1438 AAGAAGCAACAGGATCTACCTTACCTTACCTTAAAGAA-TGAGGTGGGAGTGGAATG 1496
QY 462 uValMetAlaPhePheLeuIleThrTyrPheHisSerThrTrpValThrSerGluAl 482
Db 1497 GGTATGCTTTTCTCATCACCCTACACCTTTCATTCGACTTGGGTGACCATGAGC 1556
QY 482 aTyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIlePhe 502
Db 1557 CTATTCTTCTCCCTCCATGCTGCTGCTGCTGGGAGTGGCATGAGATCAATTTT 1616
QY 502 eAspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVa 522
Db 1617 TGATGATCTCCGAGAGGCTATTATGCTCCGTCACATATCTCCAGGATGCAAAAGT 1676
QY 522 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleVal 542
Db 1677 CATGTCATGCTGGGATTTGCTACCAATTTACTGCAATGCGCAATCGCAATTTT 1736
QY 542 lAspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerTh 562
Db 1737 GGCACATAACACATGGATATATCCATATTTCTCGAGTGGGAGGATGGCATCCAC 1796
QY 562 rGluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleValIlePhe 582
Db 1797 AGAAGAAAGGCTATGAATCATGAGGAGCTTGATGTCAGCTATGCTGCTTGTCA 1856
QY 582 eGlyGlyLeuThrGlyTyrSerSerAspAspIleAsnLysPheLeuTrpMetValArgII 602
Db 1857 TGGAGGCTTACTGGGTATCTTCGGATGATATCAACAGTTTCTTTGGATGCTCGGAT 1916
QY 602 eGlyGlySerThrAsp-ThrGlyLys-HisIleLysGluAsnAspTyrTyrThrProThr 621
Db 1917 TGGAGGAACACAGA-GACAGG-AAAGACACATTAAGGAGATGACTACTATCTCTACT 1974
QY 622 GlyGluPheArgValAspArgGluGlySerProValLeuLeuAsnCysLeuMetTyrLys 641
Db 1975 GGGGATTCGGTTGATGCTGAGGGTCTCCGGTGTCTGCTCACTGCTTATGATACAA 2034
QY 642 MetCysTyrTyrArgPheGlyGlnValTyrThrGluAlaLysArgProPtoGlyPheAsp 661
Db 2035 ATGTGTACTACCGCTTTGGGAGGTCTACAGAAAGCAAGCGTCCACAGGCTTTGAC 2094
QY 662 ArgValArgAsnAlaGluIleGlyAsnLysAspPheGluLeuAspValLeuGluGlu 680

Db 2095 CGTGTTCGAAATGCTGAGATTGGTAAATAAGACTTTGAGCTTGATCTCTGGAGAA 2151

RESULT 5

ABK27622/c
ID ABK27622 standard; cDNA; 440 BP.
XX
AC ABK27622;
XX
DT 09-APR-2002 (first entry)
XX
DE Human colon cancer expressed sequence tag, Seq ID no 59.
XX
KW Human; colon cancer; T cell expansion; tumour; EST; gene; ss;
XX expressed sequence tag.

OS Homo sapiens.

XX WO200196390-A2.

XX 20-DEC-2001.

XX 08-JUN-2001; 2001WO-US018577.

XX 09-JUN-2000; 2000US-0210821P.

XX 18-DEC-2000; 2000US-0256571P.

XX 10-MAY-2001; 2001US-0290240P.

XX (CORI-) CORIYA CORP.

XX Jiang Y, Hepler WT, Clapper JD, Wang A, Secrist H;

XX WPI; 2002-139708/18.

XX Novel isolated polynucleotide encoding a polypeptide comprising a portion

XX of colon tumor protein, useful for detection, diagnosis and therapy of

XX human colon cancer.

XX Claim 1; Page 167; 220pp; English.

XX The invention relates to an isolated polynucleotide (I) encoding a

XX polypeptide (II) comprising at least a portion of a colon tumour protein.

XX (I), (II) and antibody (III) to (II) are useful for determining the

XX presence of a cancer in a patient. (I), (II) or antigen presenting cells

XX expressing (I) is useful for stimulating and/or expanding T cells

XX specific for a tumour protein, by contacting T cells with (I), (II) or

XX antigen-presenting cells that express (I), under conditions and for a

XX time sufficient to permit the stimulation and/or expansion of T cells.

XX (I), (II), or antigen presenting cells that express (I) are useful for

XX treating colon cancer in a patient by incubating CD4+ and/or CD8+ T cells

XX isolated from a patient with (I), (II) or antigen presenting cells that

XX express (II) such that T cells proliferate, and administering to the

XX patient an effective amount of the proliferated T cells, thus inhibiting

XX the development of a cancer in the patient. (I) or (II) is useful in

XX vaccines and pharmaceutical compositions for prevention and treatment of

XX colon malignancies and for the diagnosis and monitoring of such cancers.

XX (I), (II) or (III) is useful for detection, diagnosis and/or therapy of

XX human colon cancer. (I) is useful as a probe or primer for nucleic acid

XX hybridisation, and in the design and preparation of ribozyme molecules

XX for inhibiting expression of (II) in tumour cells. ABK27564-ABK27807

XX represent novel human colon cancer coding sequences and primers of the

XX invention

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Alignment Scores:
Pred. No.: 1,02e-141 Length: 440
Score: 146.00 Matches: 146
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.71% Indels: 0
DB: Gaps: 0

Sequence 440 BP; 117 A; 119 C; 84 G; 120 T; 0 U; 0 Other;

PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
 PI Catano TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 XX WPI; 2000-256641/22.
 XX
 XX Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer.
 XX
 PS Claim 16; Page 209; 345pp; English.
 XX
 CC This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridize to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage
 XX
 SQ Sequence 487 BP; 90 A; 140 C; 102 G; 155 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,75e-123 Length: 487
 Score: 128.00 Matches: 128
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 18.16% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x AAA16208 (1-487)

QY 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValIlePheValSerLysLeu 290
 Db 88 TTGGGGCTTTGGGCTCTGCGAGATCCATCCCTTTGGGATACCTGCGCAGCAGTTG 147
 QY 291 AsnProGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
 Db 148 AATCCACAAACAAATTGAAGTTCTTTTCGGAGCGTCATCTCTGTTAGGCTTTGTCTT 207
 QY 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProThrGlyArgPhe 330
 Db 208 CTCACCGTGGGAGCTCTCTCATGCTGACAGAAATAATCTCCCTGACGGGCGGTTTC 267
 QY 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnAsnIlePheIleAlaSerValSer 350
 Db 268 TACTCACTGCTGGAATCCCTCTTATGCTTAAGAACACATCCCAATGCTTCTGTGCT 327
 QY 351 GluHisGlnProThrThrTyrSerSerTyrTyrPheAspLeuGlnLeuValPheMet 370
 Db 328 GAGCATCAGCCACCAACCTGGTCTTCATACATTTTGACCTGACGCTCTGCTTCATG 387
 QY 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIle 390
 Db 388 TTTCAGTTGGCTCTTATTACTGCTTTAGCAACCTGCTGATGCCCGGATTTTATCATC 447
 QY 391 MetTyrGlyValThrSerMetTyr 398
 Db 448 ATGTATGGTGTGACCATGATGATAC 471

RESULT 8

ID ADP28508
 ADP28508 standard; DNA; 1371 BP.

XX
 AC ADP28508;

XX
 DT 12-AUG-2004 (first entry)

XX
 DE Human secreted protein encoding sequence SEQ ID #506.
 XX
 KW Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
 KW cancer; inflammatory; immune; ds; human secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO2004035732-A2.
 XX
 PD 29-APR-2004.
 XX
 XX 28-AUG-2003; 2003WO-US026780.
 XX
 XX 29-AUG-2002; 2002US-0406576P.
 XX 29-AUG-2002; 2002US-0406579P.
 XX 29-AUG-2002; 2002US-0406585P.
 XX 29-AUG-2002; 2002US-0406588P.
 XX 29-AUG-2002; 2002US-0406588P.
 XX 29-AUG-2002; 2002US-0406611P.
 XX 29-AUG-2002; 2002US-0406612P.
 XX 29-AUG-2002; 2002US-0406616P.
 XX 29-AUG-2002; 2002US-0406640P.
 XX 29-AUG-2002; 2002US-0406642P.
 XX 29-AUG-2002; 2002US-0406646P.
 XX 29-AUG-2002; 2002US-0406653P.
 XX 29-AUG-2002; 2002US-0406655P.
 XX 29-AUG-2002; 2002US-0406666P.
 XX 17-SEP-2002; 2002US-0410946P.
 XX 17-SEP-2002; 2002US-0410947P.
 XX 17-SEP-2002; 2002US-0410948P.
 XX 17-SEP-2002; 2002US-0410949P.
 XX 17-SEP-2002; 2002US-0410953P.
 XX 17-SEP-2002; 2002US-0410957P.
 XX 17-SEP-2002; 2002US-0410958P.
 XX 17-SEP-2002; 2002US-0410959P.
 XX 17-SEP-2002; 2002US-0410960P.
 XX 17-SEP-2002; 2002US-0410961P.
 XX 17-SEP-2002; 2002US-0410962P.
 XX 17-SEP-2002; 2002US-0410963P.
 XX 17-SEP-2002; 2002US-0410964P.
 XX 17-SEP-2002; 2002US-0410965P.
 XX 17-SEP-2002; 2002US-0410966P.
 XX 17-SEP-2002; 2002US-0410967P.
 XX 17-SEP-2002; 2002US-0410968P.
 XX 17-SEP-2002; 2002US-0410969P.
 XX 17-SEP-2002; 2002US-0410970P.
 XX 17-SEP-2002; 2002US-0410971P.
 XX 17-SEP-2002; 2002US-0410972P.
 XX 17-SEP-2002; 2002US-0410973P.
 XX 17-SEP-2002; 2002US-0410974P.
 XX 17-SEP-2002; 2002US-0410975P.
 XX 17-SEP-2002; 2002US-0410976P.
 XX 17-SEP-2002; 2002US-0410977P.
 XX 17-SEP-2002; 2002US-0410978P.
 XX 17-SEP-2002; 2002US-0410979P.
 XX 17-SEP-2002; 2002US-0410980P.
 XX 17-SEP-2002; 2002US-0410981P.
 XX 17-SEP-2002; 2002US-0410982P.
 XX 17-SEP-2002; 2002US-0410983P.
 XX 17-SEP-2002; 2002US-0410984P.
 XX 17-SEP-2002; 2002US-0410985P.
 XX 17-SEP-2002; 2002US-0410986P.
 XX 17-SEP-2002; 2002US-0410987P.
 XX 17-SEP-2002; 2002US-0410988P.
 XX 17-SEP-2002; 2002US-0410989P.
 XX 17-SEP-2002; 2002US-0410990P.
 XX 17-SEP-2002; 2002US-0410991P.
 XX 17-SEP-2002; 2002US-0410992P.
 XX 17-SEP-2002; 2002US-0410993P.
 XX 17-SEP-2002; 2002US-0410994P.
 XX 17-SEP-2002; 2002US-0410995P.
 XX 17-SEP-2002; 2002US-0410996P.
 XX 17-SEP-2002; 2002US-0410997P.
 XX 17-SEP-2002; 2002US-0410998P.
 XX 17-SEP-2002; 2002US-0410999P.
 XX 17-SEP-2002; 2002US-0411000P.
 XX 17-SEP-2002; 2002US-0411001P.
 XX 17-SEP-2002; 2002US-0411002P.
 XX 17-SEP-2002; 2002US-0411003P.
 XX 17-SEP-2002; 2002US-0411004P.
 XX 17-SEP-2002; 2002US-0411005P.
 XX 17-SEP-2002; 2002US-0411006P.
 XX 17-SEP-2002; 2002US-0411007P.
 XX 17-SEP-2002; 2002US-0411008P.
 XX 17-SEP-2002; 2002US-0411009P.
 XX 17-SEP-2002; 2002US-0411010P.
 XX 17-SEP-2002; 2002US-0411011P.
 XX 17-SEP-2002; 2002US-0411012P.
 XX 17-SEP-2002; 2002US-0411013P.
 XX 17-SEP-2002; 2002US-0411014P.
 XX 17-SEP-2002; 2002US-0411015P.
 XX 17-SEP-2002; 2002US-0411016P.
 XX 17-SEP-2002; 2002US-0411017P.
 XX 17-SEP-2002; 2002US-0411018P.
 XX 17-SEP-2002; 2002US-0411019P.
 XX 17-SEP-2002; 2002US-0411020P.
 XX 17-SEP-2002; 2002US-0411021P.
 XX 17-SEP-2002; 2002US-0411022P.
 XX 17-SEP-2002; 2002US-0411023P.
 XX 17-SEP-2002; 2002US-0411024P.
 XX 17-SEP-2002; 2002US-0411025P.
 XX 17-SEP-2002; 2002US-0411026P.
 XX 17-SEP-2002; 2002US-0411027P.
 XX 17-SEP-2002; 2002US-0411028P.
 XX 17-SEP-2002; 2002US-0411029P.
 XX 17-SEP-2002; 2002US-0411030P.
 XX 17-SEP-2002; 2002US-0411031P.
 XX 17-SEP-2002; 2002US-0411032P.
 XX 17-SEP-2002; 2002US-0411033P.
 XX 17-SEP-2002; 2002US-0411034P.
 XX 17-SEP-2002; 2002US-0411035P.
 XX 17-SEP-2002; 2002US-0411036P.
 XX 17-SEP-2002; 2002US-0411037P.
 XX 17-SEP-2002; 2002US-0411038P.
 XX 17-SEP-2002; 2002US-0411039P.
 XX 17-SEP-2002; 2002US-0411040P.
 XX 17-SEP-2002; 2002US-0411041P.
 XX 17-SEP-2002; 2002US-0411042P.
 XX 17-SEP-2002; 2002US-0411043P.
 XX 17-SEP-2002; 2002US-0411044P.
 XX 17-SEP-2002; 2002US-0411045P.
 XX 17-SEP-2002; 2002US-0411046P.
 XX 17-SEP-2002; 2002US-0411047P.
 XX 17-SEP-2002; 2002US-0411048P.
 XX 17-SEP-2002; 2002US-0411049P.
 XX 17-SEP-2002; 2002US-0411050P.
 XX 17-SEP-2002; 2002US-0411051P.
 XX 17-SEP-2002; 2002US-0411052P.
 XX 17-SEP-2002; 2002US-0411053P.
 XX 17-SEP-2002; 2002US-0411054P.
 XX 17-SEP-2002; 2002US-0411055P.
 XX 17-SEP-2002; 2002US-0411056P.
 XX 17-SEP-2002; 2002US-0411057P.
 XX 17-SEP-2002; 2002US-0411058P.
 XX 17-SEP-2002; 2002US-0411059P.
 XX 17-SEP-2002; 2002US-0411060P.
 XX 17-SEP-2002; 2002US-0411061P.
 XX 17-SEP-2002; 2002US-0411062P.
 XX 17-SEP-2002; 2002US-0411063P.
 XX 17-SEP-2002; 2002US-0411064P.
 XX 17-SEP-2002; 2002US-0411065P.
 XX 17-SEP-2002; 2002US-0411066P.
 XX 17-SEP-2002; 2002US-0411067P.
 XX 17-SEP-2002; 2002US-0411068P.
 XX 17-SEP-2002; 2002US-0411069P.
 XX 17-SEP-2002; 2002US-0411070P.
 XX 17-SEP-2002; 2002US-0411071P.
 XX 17-SEP-2002; 2002US-0411072P.
 XX 17-SEP-2002; 2002US-0411073P.
 XX 17-SEP-2002; 2002US-0411074P.
 XX 17-SEP-2002; 2002US-0411075P.
 XX 17-SEP-2002; 2002US-0411076P.
 XX 17-SEP-2002; 2002US-0411077P.
 XX 17-SEP-2002; 2002US-0411078P.
 XX 17-SEP-2002; 2002US-0411079P.
 XX 17-SEP-2002; 2002US-0411080P.
 XX 17-SEP-2002; 2002US-0411081P.
 XX 17-SEP-2002; 2002US-0411082P.
 XX 17-SEP-2002; 2002US-0411083P.
 XX 17-SEP-2002; 2002US-0411084P.
 XX 17-SEP-2002; 2002US-0411085P.
 XX 17-SEP-2002; 2002US-0411086P.
 XX 17-SEP-2002; 2002US-0411087P.
 XX 17-SEP-2002; 2002US-0411088P.
 XX 17-SEP-2002; 2002US-0411089P.
 XX 17-SEP-2002; 2002US-0411090P.
 XX 17-SEP-2002; 2002US-0411091P.
 XX 17-SEP-2002; 2002US-0411092P.
 XX 17-SEP-2002; 2002US-0411093P.
 XX 17-SEP-2002; 2002US-0411094P.
 XX 17-SEP-2002; 2002US-0411095P.
 XX 17-SEP-2002; 2002US-0411096P.
 XX 17-SEP-2002; 2002US-0411097P.
 XX 17-SEP-2002; 2002US-0411098P.
 XX 17-SEP-2002; 2002US-0411099P.
 XX 17-SEP-2002; 2002US-0411100P.
 XX 17-SEP-2002; 2002US-0411101P.
 XX 17-SEP-2002; 2002US-0411102P.
 XX 17-SEP-2002; 2002US-0411103P.
 XX 17-SEP-2002; 2002US-0411104P.
 XX 17-SEP-2002; 2002US-0411105P.
 XX 17-SEP-2002; 2002US-0411106P.
 XX 17-SEP-2002; 2002US-0411107P.
 XX 17-SEP-2002; 2002US-0411108P.
 XX 17-SEP-2002; 2002US-0411109P.
 XX 17-SEP-2002; 2002US-0411110P.
 XX 17-SEP-2002; 2002US-0411111P.
 XX 17-SEP-2002; 2002US-0411112P.
 XX 17-SEP-2002; 2002US-0411113P.
 XX 17-SEP-2002; 2002US-0411114P.
 XX 17-SEP-2002; 2002US-0411115P.
 XX 17-SEP-2002; 2002US-0411116P.
 XX 17-SEP-2002; 2002US-0411117P.
 XX 17-SEP-2002; 2002US-0411118P.
 XX 17-SEP-2002; 2002US-0411119P.
 XX 17-SEP-2002; 2002US-0411120P.
 XX 17-SEP-2002; 2002US-0411121P.
 XX 17-SEP-2002; 2002US-0411122P.
 XX 17-SEP-2002; 2002US-0411123P.
 XX 17-SEP-2002; 2002US-0411124P.
 XX 17-SEP-2002; 2002US-0411125P.
 XX 17-SEP-2002; 2002US-0411126P.
 XX 17-SEP-2002; 2002US-0411127P.
 XX 17-SEP-2002; 2002US-0411128P.
 XX 17-SEP-2002; 2002US-0411129P.
 XX 17-SEP-2002; 2002US-0411130P.
 XX 17-SEP-2002; 2002US-0411131P.
 XX 17-SEP-2002; 2002US-0411132P.
 XX 17-SEP-2002; 2002US-0411133P.
 XX 17-SEP-2002; 2002US-0411134P.
 XX 17-SEP-2002; 2002US-0411135P.
 XX 17-SEP-2002; 2002US-0411136P.
 XX 17-SEP-2002; 2002US-0411137P.
 XX 17-SEP-2002; 2002US-0411138P.
 XX 17-SEP-2002; 2002US-0411139P.
 XX 17-SEP-2002; 2002US-0411140P.
 XX 17-SEP-2002; 2002US-0411141P.
 XX 17-SEP-2002; 2002US-0411142P.
 XX 17-SEP-2002; 2002US-0411143P.
 XX 17-SEP-2002; 2002US-0411144P.
 XX 17-SEP-2002; 2002US-0411145P.
 XX 17-SEP-2002; 2002US-0411146P.
 XX 17-SEP-2002; 2002US-0411147P.
 XX 17-SEP-2002; 2002US-0411148P.
 XX 17-SEP-2002; 2002US-0411149P.
 XX 17-SEP-2002; 2002US-0411150P.
 XX 17-SEP-2002; 2002US-0411151P.
 XX 17-SEP-2002; 2002US-0411152P.
 XX 17-SEP-2002; 2002US-0411153P.
 XX 17-SEP-2002; 2002US-0411154P.
 XX 17-SEP-2002; 2002US-0411155P.
 XX 17-SEP-2002; 2002US-0411156P.
 XX 17-SEP-2002; 2002US-0411157P.
 XX 17-SEP-2002; 2002US-0411158P.
 XX 17-SEP-2002; 2002US-0411159P.
 XX 17-SEP-2002; 2002US-0411160P.
 XX 17-SEP-2002; 2002US-0411161P.
 XX 17-SEP-2002; 2002US-0411162P.
 XX 17-SEP-2002; 2002US-0411163P.
 XX 17-SEP-2002; 2002US-0411164P.
 XX 17-SEP-2002; 2002US-0411165P.
 XX 17-SEP-2002; 2002US-0411166P.
 XX 17-SEP-2002; 2002US-0411167P.
 XX 17-SEP-2002; 2002US-0411168P.
 XX 17-SEP-2002; 2002US-0411169P.
 XX 17-SEP-2002; 2002US-0411170P.
 XX 17-SEP-2002; 2002US-0411171P.
 XX 17-SEP-2002; 2002US-0411172P.
 XX 17-SEP-2002; 2002US-0411173P.
 XX 17-SEP-2002; 2002US-0411174P.
 XX 17-SEP-2002; 2002US-0411175P.
 XX 17-SEP-2002; 2002US-0411176P.
 XX 17-SEP-2002; 2002US-0411177P.
 XX 17-SEP-2002; 2002US-0411178P.
 XX 17-SEP-2002; 2002US-0411179P.
 XX 17-SEP-2002; 2002US-0411180P.
 XX 17-SEP-2002; 2002US-0411181P.
 XX 17-SEP-2002; 2002US-0411182P.
 XX 17-SEP-2002; 2002US-0411183P.
 XX 17-SEP-2002; 2002US-0411184P.
 XX 17-SEP-2002; 2002US-0411185P.
 XX 17-SEP-2002; 2002US-0411186P.
 XX 17-SEP-2002; 2002US-0411187P.
 XX 17-SEP-2002; 2002US-0411188P.
 XX 17-SEP-2002; 2002US-0411189P.
 XX 17-SEP-2002; 2002US-0411190P.
 XX 17-SEP-2002; 2002US-0411191P.
 XX 17-SEP-2002; 2002US-0411192P.
 XX 17-SEP-2002; 2002US-0411193P.
 XX 17-SEP-2002; 2002US-0411194P.
 XX 17-SEP-2002; 2002US-0411195P.
 XX 17-SEP-2002; 2002US-0411196P.
 XX 17-SEP-2002; 2002US-0411197P.
 XX 17-SEP-2002; 2002US-0411198P.
 XX 17-SEP-2002; 2002US-0411199P.
 XX 17-SEP-2002; 2002US-0411200P.
 XX 17-SEP-2002; 2002US-0411201P.
 XX 17-SEP-2002; 2002US-0411202P.
 XX 17-SEP-2002; 2002US-0411203P.
 XX 17-SEP-2002; 2002US-0411204P.
 XX 17-SEP-2002; 2002US-0411205P.
 XX 17-SEP-2002; 2002US-0411206P.
 XX 17-SEP-2002; 2002US-0411207P.
 XX 17-SEP-2002; 2002US-0411208P.
 XX 17-SEP-2002; 2002US-0411209P.
 XX 17-SEP-2002; 2002US-0411210P.
 XX 17-SEP-2002; 2002US-0411211P.
 XX 17-SEP-2002; 2002US-0411212P.
 XX 17-SEP-2002; 2002US-0411213P.
 XX 17-SEP-2002; 2002US-0411214P.
 XX 17-SEP-2002; 2002US-0411215P.
 XX 17-SEP-2002; 2002US-0411216P.
 XX 17-SEP-2002; 2002US-0411217P.
 XX 17-SEP-2002; 2002US-0411218P.
 XX 17-SEP-2002; 2002US-0411219P.
 XX 17-SEP-2002; 2002US-0411220P.
 XX 17-SEP-2002; 2002US-0411221P.
 XX 17-SEP-2002; 2002US-0411222P.
 XX 17-SEP-2002; 2002US-0411223P.
 XX 17-SEP-2002; 2002US-0411224P.
 XX 17-SEP-2002; 2002US-0411225P.
 XX 17-SEP-2002; 2002US-0411226P.
 XX 17-SEP-2002; 2002US-0411227P.
 XX 17-SEP-2002; 2002US-0411228P.
 XX 17-SEP-2002; 2002US-0411229P.
 XX 17-SEP-2002; 2002US-0411230P.
 XX 17-SEP-2002; 2002US-0411231P.
 XX 17-SEP-2002; 2002US-0411232P.
 XX 17-SEP-2002; 2002US-0411233P.
 XX 17-SEP-2002; 2002US-0411234P.
 XX 17-SEP-2002; 2002US-0411235P.
 XX 17-SEP-2002; 2002US-0411236P.
 XX 17-SEP-2002; 2002US-0411237P.
 XX 17-SEP-2002; 2002US-0411238P.
 XX 17-SEP-2002; 2002US-0411239P.
 XX 17-SEP-2002; 2002US-0411240P.
 XX 17-SEP-2002; 2002US-0411241P.
 XX 17-SEP-2002; 2002US-0411242P.
 XX 17-SEP-2002; 2002US-0411243P.
 XX 17-SEP-2002; 2002US-0411244P.
 XX 17-SEP-2002; 2002US-0411245P.
 XX 17-SEP-2002; 2002US-0411246P.
 XX 17-SEP-2002; 2002US-0411247P.
 XX 17-SEP-2002; 2002US-0411248P.
 XX 17-SEP-2002; 2002US-0411249P.
 XX 17-SEP-2002; 2002US-0411250P.
 XX 17-SEP-2002; 2002US-0411251P.
 XX 17-SEP-2002; 2002US-0411252P.
 XX 17-SEP-2002; 2002US-0411253P.
 XX 17-SEP-2002; 2002US-0411254P.
 XX 17-SEP-2002; 2002US-0411255P.
 XX 17-SEP-2002; 2002US-0411256P.
 XX 17-SEP-2002; 2002US-0411257P.
 XX 17-SEP-2002; 2002US-0411258P.
 XX 17-SEP-2002; 2002US-0411259P.
 XX 17-SEP-2002; 2002US-0411260P.
 XX 17-SEP-2002; 2002US-0411261P.
 XX 17-SEP-2002; 2002US-0411262P.
 XX 17-SEP-2002; 2002US-0411263P.
 XX 17-SEP-2002; 2002US-0411264P.
 XX 17-SEP-2002; 2002US-0411265P.
 XX 17-SEP-2002;

PR 08-JUL-2003; 2003US-0485325P.
 PR 14-JUL-2003; 2003US-0486446P.
 PR 14-JUL-2003; 2003US-0486480P.
 PR 15-JUL-2003; 2003US-0486891P.
 PR 08-AUG-2003; 2003US-0486960P.
 PR 08-AUG-2003; 2003US-0493341P.
 PR 08-AUG-2003; 2003US-0493370P.
 PR 08-AUG-2003; 2003US-0493573P.
 PR 08-AUG-2003; 2003US-0493577P.
 XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
 XX
 PI Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
 PI Halenbeck RF, Huang MM, Kothakota S, Haishan L, Linnemann T;
 PI Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
 XX
 DR WPI, 2004-348438/32.
 XX
 XX New nucleic acid molecule for diagnosing, preventing or treating diseases
 PT such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
 PT genetic, bacterial and viral diseases.
 XX
 PS Claim 1; SEQ ID NO 506; 428bp; English.
 XX
 CC The present invention relates to an isolated nucleic acid molecule
 CC encoding a polypeptide which is believed to be cytostatic,
 CC antiinflammatory, immunosuppressive, antibacterial and virucidal. The
 CC composition and methods are useful for diagnosing, preventing and
 CC treating diseases such as proliferative (e.g. cancer), inflammatory,
 CC immune, metabolic, genetic, bacterial and viral diseases. The present
 CC sequence represents a human secreted protein encoding sequence. The
 CC present sequence is available on WIPONEB and is not in the specification.
 XX
 SQ Sequence 1371 BP; 283 A; 350 C; 298 G; 440 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.69e-121 Length: 1371
 Score: 127.00 Matches: 408
 Percent Similarity: 97.61% Conservative: 0
 Best Local Similarity: 97.61% Mismatches: 6
 Query Match: 18.01% Indels: 10
 DB: 12 Gaps: 0

US-10-028-384-12 (1-705) x ADP28508 (1-1371)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGluLysGlnAspThrLeuLeuLysLeu 20
 DB 1 ATGACTAAGTTTGGATTTTGGCATTTGCTATGAGACGAGGACACACTTTTGAAGCTT 60
 QY 21 LeuLeuLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 61 CTCATTCTGCAATGGCTGCTGATTTATTCCTTCTCCACTCGTCTGTGCTGCTGAGA 120
 QY 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 DB 121 TTTCAAAGTGTATCCATGATGTTGATTCCTACTTTAATATCGGACTACACAGGTTCTG 180
 QY 51 AlaGluGluGlyPheThrLysPheHisAsnTrpPheAspAspArgAlaTrpTyrProLeu 80
 DB 181 GCTGAGGAGGGGTTTATAAATCCCAATGCTGTTTATGATGACCGAGCCTGSPACCTTTG 240
 QY 81 GlyArgIleIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 DB 241 GGACGAATATTGGAGGACAAATTTACCCAGGTTTAAATCATCACCTCTGCTGCAATCTAC 300
 QY 101 HisValLeuHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 DB 301 CATGTACTCCATTTTTCACATCACCACATCGACATTCGGAATGTCTGTGTGTCTCTGCC 360
 QY 121 ProLeuPheSerSerPheThrSer-IleValThrTyrLeu-LeuThrLysGluLeuLysA 140
 DB 361 CCTCTCTTCT 418

QY 140 splaGlyAlaGlyLeuLeuAlaAlaMetIleAlaValValProGlyTyrIleSerA 160
 DB 419 ATGCAGGGCTGGGCTTCTTGTCTGCTGCATGATTGCTGTAGTTCCTGGATATATCTCCC 478
 QY 160 xgSerValAlaGlySerTyrAspAsnGluGlyIleAlaIlePheCysMetLeuLeuThr 180
 DB 479 GATCTGTGGCTGGCTCTCTATGATATGAAGGATTGCCATCTTTTGATGCTACTCACT 538
 QY 180 yTrMetTrpIleLysAlaValLysThrGlySerIleCysTrpAlaAlaLysCysAla 200
 DB 539 ACTACATGGATCAAGGAGGAGTGGTTCATCTGTGGGAGCTAAAGTGTGCC 598
 QY 200 euAlaTyrPheTyrMetValSerSerTyrGlyGlyTyrValPheLeuIleAsnLeuIleP 220
 DB 599 TTGCTTATTTTACATGGTCTGTCATGGGAGGTTATGTCTCTGATCACTTAATTC 658
 QY 220 roleuHisValLeuValLeuMetLeuThrGlyArgPheSerHisArgIleTyrValAla 240
 DB 659 CTCCTCACGCTCTGCTGCTGATGCTCACAGGCGGTTCTCTCACCGGATCTATGTGCC 718
 QY 240 yrcysThrValTyrCysLeuGlyThrIleLeuSerArg-GlnIleSerPheValGlyPhe 259
 DB 719 ACTGTACTCTTACTGCTGGGCACTATATCTTCTAT-GCAGATCTCTCTTGTGGTTTC 777
 QY 260 GlnProValLeuSerSerGlnHisMetAlaGly-PheGlyValPheGlyLeuCysGlnI 279
 DB 778 CAGCGTGTCTCTTCTATCAGAGCACATGGCAGC-CTTTGGGGTCTTTGGTCTCTGCCAG 836
 QY 279 eHisAlaPheValAspTyrLeuArgSerLysLeuAsnProGlnGlnPheGluValLeuPh 299
 DB 837 CCATGCTTTTGGATTACCTTGGCAGCAAGTTGAATCCACCAACAATTTGAAGTCTTTT 896
 QY 299 eArgSerValIleSerLeuValGlyPheValLeuLeuThrValGlyAlaLeuMetLe 319
 DB 897 CCGGAGCGCATCTCTCTGGTAGGCTTTGTCTCTCACGCTGGAGCTCTCTCATGC- 955
 QY 319 uThrGlyLysIleSerProTyrThrGlyArgPheTyr-SerLeuLeuAspProSerTyrAl 339
 DB 956 -TG-GGAAAAATATCTCTCGAGCGGGCGTTTCTACTCGCTGCTGCATCCCTTATGC 1013
 QY 339 alyAsnAsnIleProIleIleAlaSerValSerGluHisGlnProThrThrTrpSerSe 359
 DB 1014 TAAGAACACATCCCATCATTTGCTTCTGTCTGAGCATCAGCCCAACACCTGGTCTC 1073
 QY 359 rTyTyrPheAspLeuGlnLeuLeuValPheMetPheProValGlyLeuTyrTyCysPh 379
 DB 1074 ATACTATTTGACCTGACCTCTCTCTCATGTTTCCAGTTGGCTCTATTACTGCTT 1133
 QY 379 eSerAsnLeuSerAspAlaArgIlePheIleIleMetTyrGlyValThrSerMetTyrPh 399
 DB 1134 TAGCAACCTGTCTGATGCCCGGATTTTATCATCATGATGTTGGTGACCGCATGTACTT 1193
 QY 399 eSerAlaValMetValArgLeuMetLeuValLeuAlaProValMet 414
 DB 1194 TTCAGCTGTAATGGTGGCTGCTAAAGTGTAGTGTGGCACTGTATTG 1239

RESULT 9
 AAC98825
 ID AAC98825 standard; cDNA; 575 BP.
 XX
 AC AAC98825;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:53.
 XX
 KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
 KW diagnosis; identification; cytostatic; neuroprotective; nootropic;
 KW immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW immune system; muscular; reproductive; gastrointestinal; pulmonary;
 KW cardiovascular; renal; proliferative; ss.

XX OS Homo sapiens.
 XX PN WO200055320-A1.
 XX PD 21-SEP-2000.
 XX PF 08-NAR-2000; 2000WO-US005989.
 XX PR 12-NAR-1999; 99US-0124270P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM;
 XX DR WPI; 2000-579444/54.
 XX DR P-PSDB; AAB54060.
 XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition.
 XX PS Claim 1; Page 533; 1379pp; English.
 XX CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nocotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention
 XX SQ Sequence 575 BP; 117 A; 144 C; 137 G; 174 T; 0 U; 3 Other;

Alignment Scores:
 Pred. No.: 8.08e-121 Length: 575
 Score: 126.00 Matches: 126
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.87% Indels: 0
 DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x AAC98825 (1-575)

QY 1 MetThrLysPheGlyPheLeuArgLeuSerTyGluLysClnAspThrLeuLeuLysLeu 20
 DB 115 ATGACTAAGTTTGATTTTGGCATTGTCTATGAGAACGAGACACACTTTTGAAGCTT 174
 QY 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
 DB 175 CTCATTCGTGTCATGGCTGCTGATTATATCTCTCCACTCGCTGTTCCTGCTCGAGA 234
 QY 41 PheGluSerValIleHisGluPheAspProTyPheAsnTyArgThrThrArgPheLeu 60
 DB 235 TTTGAAAGTGTTATCCATGAGTTTGATCCGCTATTAAATTCGGACTACCAAGTTCCTG 294
 QY 61 AlaGluGluGlyPheTyTrLysPheHisAsnTrpPheAspAspAlaTrpTyProLeu 80

Db 295 GCTGAGGAGGGGTTTATATAATTCATAACTGTTTGTATGACGACGACCTGGTACCTTTG 354
 QY 81 GlyArgIleIleGlyGlyThrIleTyProGlyLeuMetIleThrSerAlaIleTy 100
 Db 355 GGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAAATGATCACCTCTGTCGAATCTAC 414
 QY 101 HisValIleHisPhePheHisIleThrIleAspIleArgAsnValCysValPheLeuAla 120
 Db 415 CATGTACTCCATTTTTCACATCCACATCCGACATTCGGAATGTCTGTGTCTCTGGCC 474
 QY 121 ProLeuPheSerSerPhe 126
 Db 475 CCTCTCTCTCTCTCTTC 492

RESULT 10
 AAC00273
 ID AAC00273 standard; cDNA; 414 BP.
 XX AC AAC00273;
 XX DT 06-OCT-2000 (first entry)
 XX DE Human secreted protein 5' EST, SEQ ID NO: 271.
 XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 XX gene therapy; chromosome mapping; ss.
 XX OS Homo sapiens.
 XX PN EP1033401-A2.
 XX PD 06-SEP-2000.
 XX PF 21-FEB-2000; 2000BP-00200610.
 XX PR 26-FEB-1999; 99US-0122487P.
 XX PA (GEST) GENSET.
 XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.
 XX DR P-PSDB; AAG00267.
 XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX PS Claim 1; SEQ ID NO 271; 71pp + Sequence Listing; English.
 XX CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors
 XX SQ Sequence 414 BP; 89 A; 93 C; 105 G; 127 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5.22e-96 Length: 414
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0

CC for determining the presence of a cancer in a patient. A composition
CC containing the polynucleotide and/or polypeptide is useful for treating a
CC lung cancer in a patient. The polypeptide is useful for removing tumour
CC cells from a biological sample. The polynucleotide is also useful as
CC probe or primer to detect the level of mRNA encoding a tumour protein.
CC This sequence encodes a lung tumour associated protein or protein
CC fragment, described in the method of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC
XX Sequence 487 BP; 116 A; 115 C; 135 G; 121 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 6.09e-96 Length: 487
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.47% Indels: 0
DB: 6 Gaps: 0

US-10-028-384-12 (1-705) x ABX39617 (1-487)
Qy 483 TyrSerSerProSerIleValIleuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 77 TACTCTCTCCGTCCATTTACTATCTGCTCCGTGGGGATGGCAGTAGGATCATATTT 136
Qy 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaIleVal 522
Db 137 GATGACTTCGAGAGCATATTTATTTGCTCCGTCTATACTCCAGAGATCGAAGGTC 196
Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 197 ATGCTCTGGTGGGATTATGCTATCATGATACAGCTATGCGAACCAGAACAAATTTAGG 256
Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 257 GACATACACATGAGTATATACCATATTTCTCGAGTAGGCGAGCAATGGGTCCACA 316
Qy 563 GluGluAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleValIlePhe 582
Db 317 GAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT 376
Qy 583 GlyGly 584
Db 377 GGAGGA 382

RESULT 12
ACAl1946
ID ACAl1946 standard; cDNA; 487 BP.
XX
AC ACAl1946;
XX
DT 06-JUN-2003 (first entry)
XX
DE Human lung neuroendocrine carcinoma library, cDNA SEQ ID 1655.
XX Human; lung cancer; ss; lung tumour; cytostatic; vaccine;
KW T cell expansion; CD4; CD8.
XX
OS Homo sapiens.
XX
PN US2002197669-A1.
XX
PD 26-DEC-2002.
XX
PF 03-MAY-2001; 2001US-00849626.
XX
PR 13-DEC-2000; 2000US-00736457.
XX
PA (BANG/) BANGUR C S.
PA (FANG/) FANGER G R.
PA (WANG/) WANG A.

DB: 3 Gaps: 0
US-10-028-384-12 (1-705) x AAC00273 (1-414)
Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
Db 108 ATGACTAAGTTGGATTTTGGATTGTCCTATGGAAGCAGGACACATTTTGAAGCTT 167
Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40
Db 168 CTCATTCCTCAATGGCTGCTATTATCTCTCCACTCGTCTGTTGCTGCTCGAGA 227
Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrArgPheLeu 60
Db 228 TTGAAAGTGTATCATGAGTTTGCCTACTTAAATATCGGACTACACAGTTCTCTG 287
Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTrpPheAspArgAlaTyrTrpLeu 80
Db 288 GCTGAGGAGGGTTTATTAATCCATAACTGTTTGTATGACCGAGCCTGGTACCCCTTG 347
Qy 81 GlyArgIleIleGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
Db 348 GGACGAATCATTTGGAGAACAAATTTACCAGGTTTAAATGATCACCTCTGCTGCAATCTAC 407
Qy 101 HisVal 102
Db 408 CATGTA 413

RESULT 11
ABX39617
ID ABX39617 standard; cDNA; 487 BP.
XX
AC ABX39617;
XX
DT 21-MAY-2002 (first entry)
XX
DE cDNA encoding lung tumour protein clone 26496.
XX
KW Lung tumour; cancer; T cell; immune response stimulator; cytostatic;
XX Gene; ss.
OS Homo sapiens.
PN WO200204514-A2.
XX
PD 17-JAN-2002.
XX
PF 10-JUL-2001; 2001WO-US022058.
XX
PR 11-JUL-2000; 2000US-00614124.
PR 29-AUG-2000; 2000US-00651563.
PR 08-SEP-2000; 2000US-00658824.
PR 26-SEP-2000; 2000US-00671325.
PR 06-OCT-2000; 2000US-00677419.
PR 30-OCT-2000; 2000US-00702705.
PR 13-DEC-2000; 2000US-00736457.
PR 03-MAY-2001; 2001US-00849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;
PI Wang A, Fanger N, Switzer A, McNeill PD, Clapper JB;
XX
DR WPI; 2602-164634/21.
XX
PT Novel polynucleotide encoding a lung tumor polypeptide useful for
PT stimulating and/or expanding T cells specific for a tumor protein.
XX
XX
PS Claim 1; SEQ ID NO 1655; 223pp; English.
XX
XX The invention describes an isolated polynucleotide and polypeptide useful
CC for stimulating and/or expanding T cells specific for a tumour protein

PA (WANG//) WANG T.
 PA (SWIT//) SWITZER A P.
 PA (MCNE//) MCNEILL P D.
 PA (CLAP//) CLAPPER J D.
 XX
 PI Bangur CS, Fanger GR, Wang A, Wang T, Switzer AP, McNeill PD;
 PI Clapper JD;
 XX
 DR WPI; 2003-352750/33.
 XX
 XX Novel lung cancer polynucleotide encoding lung cancer protein, useful for
 PT detecting the presence of lung cancer in a patient, and in pharmaceutical
 PT compositions, e.g. vaccines, for treating lung cancer.
 XX
 PS Example 1; Page; 72pp; English.
 XX
 CC The invention relates to a polynucleotide encoding a lung tumour protein,
 CC comprising a sequence selected from any of the 14 sequences mentioned in
 CC the specification, or a sequence (S2) mentioned in specification,
 CC complement of S1, sequences consisting of at least 20 contiguous residues
 CC of S1, sequences that hybridise to S1, sequences having 75%, preferably
 CC 90%, identity to S1, or degenerate variants of S1. Also included are an
 CC isolated polypeptide (comprising a sequence (S3) selected from any one of
 CC the 4 amino acid sequences mentioned in the specification, a sequence
 CC encoded by the polynucleotide, or sequences having at least 70%
 CC preferably 90%, identity to a sequence encoded by the polynucleotide), an
 CC expression vector comprising the polynucleotide operably linked to an
 CC expression control sequence, a host cell transformed or transfected with
 CC the vector, an isolated antibody (or its antigen-binding fragment) that
 CC specifically binds to the polypeptide, detecting the presence of a cancer
 CC in a patient, a fusion protein comprising the polypeptide, an
 CC oligonucleotide that hybridises to S1 under moderately stringent
 CC conditions, stimulating and/or expanding T cells specific for a tumour
 CC protein (comprising contacting T cells with the polynucleotide, protein
 CC or antigen-presenting cells, under conditions and for a time sufficient
 CC to permit the stimulation and/or expansion of T cells) and inhibiting the
 CC development of a cancer in a patient (by incubating CD4⁺ and/or CD8⁺ T
 CC cells isolated from a patient with the polynucleotide, protein or antigen
 CC presenting cells that express the polynucleotide, such that T cells
 CC proliferate, administering to the patient an effective amount of the
 CC proliferated T cells, and thus inhibiting the development of a cancer in
 CC the patient. The polynucleotide, protein and cells are useful in a
 CC composition for stimulating an immune response in a patient, and for
 CC treating a cancer in a patient (particularly lung cancer). The
 CC oligonucleotide is useful for determining the presence of a cancer in a
 CC patient. The protein and oligonucleotides are useful in pharmaceutical
 CC compositions, e.g. vaccines. The polynucleotide is also useful as a probe
 CC or primer for nucleic acid hybridisation, and in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour
 CC polypeptides and proteins in tumour cells. An amplified portion of the
 CC polynucleotide is useful for isolating a full-length gene from a suitable
 CC library. The present sequence is a cDNA (full length, extended or
 CC partial) isolated from a library derived from lung tumour/cancer cells.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO at seqdata.uspto.gov/sequence.html?docId=20020197669
 XX
 SQ Sequence 487 BP; 116 A; 115 C; 135 G; 121 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 6.09e-96 Length: 487
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0
 DB: 8 Gaps: 0
 US-10-028-384-12 (1-705) x ACA11946 (1-487)
 QY 483 TyrSerSerProSerIleValIeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
 Db 77 TACTCTTCCTCCGTCATTGCTACTATCTGCCCTGGTGGGAGTGGCAGTAGGATCATATTT 136

QY 503 AspAspPheArgGluAlaTyrTyrTrpIleuArgHisAsnThrProGluAspAlaIysVal 522
 Db 137 GATGACTTCGAGAGAGCATATTATTGGCTCCGTCATATACTCCAGAGGATCCGAAGGTC 196
 QY 523 MetSerTyrTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 197 ATGTCTCGTGGGATTATGGCTATCAGATTACAGCTATGGCAACCGAACATTTTAGTG 256
 QY 543 AspAsnAsnThrTyrAsnAsnThrHisIleSerArgValGlyGlnAlaValAlaSerThr 562
 Db 257 GACAATAACACATGAAGTAATAATACCATATTTTCGAGTAGGCGAGCAATGGCGTCCACA 316
 QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValIleuValIlePhe 582
 Db 317 GAGGAAAAGCCATGAGATCATGAGGAGGCTCGATGTCAGCTATGCTGGTCAATTTT 376
 QY 583 GlyGly 584
 Db 377 GGAGGA 382

RESULT 13
 ACA03132
 ID ACA03132 standard; cDNA; 487 BP.
 XX ACA03132;
 AC ACA03132;
 DT 22-MAY-2003 (first entry)
 XX
 DE Lung cancer therapy and diagnosis associated cDNA #1621.
 XX
 KW Lung cancer; cytostatic; vaccine; gene therapy; cancer; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN US2002172952-A1.
 XX
 PD 21-NOV-2002.
 XX
 PF 10-JUL-2001; 2001US-00902941.
 XX
 PR 30-JUN-1999; 99US-00346492.
 PR 15-OCT-1999; 99US-00419356.
 PR 17-DEC-1999; 99US-00466867.
 PR 30-DEC-1999; 99US-00476300.
 PR 06-MAR-2000; 2000US-00519642.
 PR 22-MAR-2000; 2000US-00533077.
 PR 10-APR-2000; 2000US-00546259.
 PR 27-APR-2000; 2000US-00560406.
 PR 05-JUN-2000; 2000US-00589184.
 PR 11-JUL-2000; 2000US-00614124.
 PR 29-AUG-2000; 2000US-00651563.
 PR 08-SEP-2000; 2000US-00658824.
 PR 26-SEP-2000; 2000US-00671325.
 PR 06-OCT-2000; 2000US-00677419.
 PR 30-OCT-2000; 2000US-00702705.
 PR 13-DEC-2000; 2000US-00736457.
 PR 03-MAY-2001; 2001US-00849626.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Henderson RA, Wang T, Watanabe Y, Johnson JC, Retter MW;
 PI Durham M, Carter D, Fanger GR, Vedvick TS, Bangur CS, McNabb A;
 XX WPI; 2003-328427/31.
 XX
 XX New polynucleotide, useful for preparing a composition for treating or
 PT inhibiting development of cancer, e.g. lung cancer.
 XX
 PS Example 1; SEQ ID NO 1655; 82pp; English.
 XX
 CC The invention describes an isolated polynucleotide comprising one of 32
 CC sequences, complement or degenerate variants of them. The polynucleotide
 CC is useful for preparing a composition e.g. a vaccine or for gene therapy.

CC for treating or inhibiting development of cancer, e.g. lung cancer. This
CC sequence represents a polynucleotide associated with the compositions and
CC methods for the therapy and diagnosis of lung cancer

Sequence 487 BP; 116 A; 115 C; 135 G; 121 T; 0 U; 0 Other;

| | | |
|------------------------|----------|---------------|
| Alignment Scores: | | |
| Pred. No.: | 6,098-96 | Length: |
| Score: | 102.00 | Matches: |
| Percent similarity: | 100.00% | Conservative: |
| Best local similarity: | 100.00% | Mismatches: |
| Query Match: | 14.4% | Indels: |
| DB: | 8 | Gaps: |

UUS-10-028-384-12 (1-705) x ACA03132 (1-487)

| | | | |
|-----|----|---|-----|
| 483 | QY | TyrSerSerProSerIleValLeuSerAlaArgGlyAlaAspGlySerArgIleLeuPhe | 502 |
| 77 | DB | TACTCTTCGCTCCATTGACTACTCTGCCGCTGGTGCGATGGCAGTAGGATCATATT | 136 |
| 503 | QY | AspAspPheArgGluAlaIleTyrTyrTrpLeuArgHisAsnThrProGluAspAlaIleVal | 522 |
| 137 | DB | GATGACTTCGAGGAAGCATATTATTGGCTCGGTCATATATCTCCAGAGGATGCGAAGTTC | 196 |
| 523 | QY | MetSerTrpTrpAspTyrGlyTyrGluIleThrAlaMetAlaAsnArgThrIleLeuVal | 542 |
| 197 | DB | ATGTCCTCGTGGGATTATGGCTATACAGATTACAGCTATGGCAACCGAACTTTTAGTG | 256 |
| 543 | QY | AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr | 562 |
| 257 | DB | GACAATAACACATCGAATAATACCCATATTCTCGAGTAGGGCAGCAATGGGTCCACA | 316 |
| 563 | QY | GluGluTyrAlaIleTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe | 582 |
| 317 | DB | GAGGAAAAGCGCTATGAGATCATGAGGAGAGCTCGATGTCAGCTATGCTGTGTCATT | 376 |
| 583 | QY | GlyGly | 584 |
| 377 | DB | GGAGGA | 382 |

RESULT 14

ADH47174
ID ADH47174 standard; cDNA: 487 BP.

AC ADH47174:

25-MAR-2004 (first entry)

Human lung tumour cDNA clone SEQ ID NO 1655

XX lung cancer; tumour; immunostimulant; vaccine; gene therapy;
KW human; clone. ss

[illegible]

XX
DN

XX
09-MAY-2003[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PR 28-MAR-2002; 2002US-00113872.

PA (CORI-) CORIXA CORP.

PI Henderson RA, Wang T, Watanabe Y, Kalos MD, Sleath PR;
PI Johnson JC, Retter MW, Durham M, Carter D, Fanger GR, Vedrick TS;
PI Bangor CS, Mcnabb A;

DR WPI; 2003-468346/44.

PT New polypeptides and encoding polynucleotides, useful for diagnosing,

PT preventing and/or treating lung cancer.
XX
PS Example 1: SEO ID NO 1655: 258pp; English.

The invention relates to novel compositions and methods for the therapy and diagnosis of cancer, particularly lung cancer. The compositions comprise one or more lung tumour polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cells that express such polypeptides, and T cells that are specific for cells expressing such polypeptides. The novel compositions have cytostatic and immunostimulant activity. The lung tumour antigens can be used in the creation of a vaccine. The polynucleotides that encode the lung tumour polypeptides can be used in gene therapy to help in the treatment of lung tumours. This polynucleotide sequence represents a human lung tumour cDNA clone of the invention. This sequence was not shown in the specification. It has been taken from a World Intellectual Property Organization CD ROM supplied with the specification.

Sequence 487 BP: 116 A: 115 C: 135 G: 121 T: 0 U: 0 Other: 0

| Alignment Scores: | |
|------------------------|----------|
| Pred. NO.: | 6.09e-96 |
| Score: | 102.00 |
| Percent Similarity: | 100.00% |
| Best Local Similarity: | 100.00% |
| Query Match: | 14.47% |
| Da: | 10 |
| Indels: | 0 |
| Mismatches: | 0 |
| Conservative: | 0 |
| Matches: | 102 |
| Length: | 487 |

UIS-10-028-384-12 (1-205) x ADH47174 (1-487)

| | | |
|-----|---|-----|
| 483 | TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe | 502 |
| 77 | TACTCTTCTCCGTCCATTGACTATCTGCCGTGGCGATGGCAGTAGGCATCATATT | 136 |
| 503 | AspAspPheArgGlualatYrTrpLeuA-gHisAsnThrProGluAspAlaIysVal | 522 |
| 137 | GATGACTCCGAGAAGCATATTATTGGCTCCGTATAATACCTCCAGAGGATCGGAAGTC | 196 |
| 523 | MetSerTrpTrpAspTYrGlyTYrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal | 542 |
| 197 | ATGTCTTGTTGGGATTATGGCTATCAGATTACAGTATGGCAACCAGAACTTTTATGTG | 256 |
| 543 | AspaenAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr | 562 |
| 257 | GACAATAACCATGGGAATATATCCCATATTCTCGAGTAGGCGAGCGAATGGCGTCCACA | 316 |
| 563 | GluGluIysAlatYrGlnIleMetArgGluLeuAspValSerTyrsValLeuValIlePhe | 582 |
| 317 | GAGCAAAAAGCCATGACAGATCATGAGGGAGCTCGATGCAGCTATGTCTGTCGTCATT | 376 |
| 583 | GlyGly | 584 |

D6

RESULT 15

ID AAF68440

AC AAF68440;

DT 12-APR-2001 (first entry)
XX
DE Human lung tumour: protein related nucleotide sequence SEQ ID NO:375.

Human; lung ca

KW cytostatic; ant
XX
OS Homo sapiens.

04-JAN-2001.

```
XX PF 30-JUN-2000; 2000WO-US018061.
XX PR 30-JUN-1999; 99US-00346492.
XX PR 15-OCT-1999; 99US-00419356.
XX PR 17-DEC-1999; 99US-00466867.
XX PR 30-DEC-1999; 99US-00476300.
XX PR 06-MAR-2000; 2000US-00519642.
XX PR 22-MAR-2000; 2000US-00533077.
XX PR 10-APR-2000; 2000US-00546259.
XX PR 27-APR-2000; 2000US-00560406.
XX PR 05-JUN-2000; 2000US-00589184.
XX PA (CORI-) CORIXA CORP.
XX PI Wang T, Bangur CS, Lodes MJ, Fangner GR, Vedvick TS, Carter D;
XX PI Retter MW, Mannion J;
XX DR WPI; 2001-071488/08.
XX XX
XX XX Lung tumor-associated proteins and the nucleic acids that encode them,
XX PT useful for preventing, diagnosing and treating lung cancer.
XX XX
XX PS Claim 4; Page 282; 436pp; English.
XX XX
XX CC The present invention describes immunogenic portions of lung tumour-
XX CC associated proteins (I) and the nucleic acids (NAs) that encode them. (I)
XX CC have cytostatic activity and can be used in gene therapy, antisense
XX CC inhibition and in vaccines. The NAs and the lung tumour-associated
XX CC proteins they encode may be used in the prevention, treatment and
XX CC diagnosis of diseases associated with their inappropriate expression,
XX CC especially lung cancers. For example, the NAs may be administered to
XX CC treat diseases by rectifying mutations or deletions in a patient's genome
XX CC that affect the activity of the protein by expressing inactive proteins
XX CC or to supplement the patients own production of (I). Additionally, the
XX CC NAs may be used to produce the lung-tumour associated protein, according
XX CC to standard recombinant DNA methodology. Conversely, antisense NA
XX CC molecules may be administered to down regulate protein expression by
XX CC binding with the cells own genes and preventing their expression. The NA
XX CC and complementary sequences may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar NA sequences in
XX CC samples, and hence which patients may be in need of treatment for lung
XX CC cancer. The (I) may be used as antigens in the production of antibodies
XX CC and in assays to identify modulators (agonists and antagonists) of the
XX CC expression and activity of the protein. AAF68083 to AAF68878 and AAF76848
XX CC to AAF76878 represent human lung tumour protein related nucleotide and
XX CC protein sequences which are used in the exemplification of the present
XX CC invention
XX SQ Sequence 307 BP; 82 A; 62 C; 80 G; 83 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.3e-95 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: Gaps: 0

US-10-028-384-12 (1-705) x AAF68440 (1-307)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCCGTCCTATCTGACCTATCTGCGCGGTGGTGGGATGGCAGTAGGATCATATTT 62

QY 503 AspAspPheArgGluAlaTyrrTyrTrpLeuArgHisAsnThrProGluAspAlaIysVal 522
Db 63 GATGACTTCGAGAAGCATATTATTGGCTCCGTCATAATACTCCAGAGGATCGAAGGTC 122

QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 123 ATGTCTCTGTGGATTATGCTATCAGATTACAGTATGCGCAACCGAACAAATTTTAGTG 182
```

```
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACAATAACACATGGAAATAATACCCATATTTCTCGAGTAGGCGAGCAATGGCGTCCACA 242

QY 563 GluGluLysAlaIleTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAAGCCCTATGAGATCATGAGGAGGCTCGATGTGAGCTATGTGCTGGTCAATTTT 302

QY 583 Gly 583
Db 303 GGA 305
```

Search completed: December 14, 2004, 23:35:26
Job time : 836.278 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 14, 2004, 13:16:57 ; Search time 137.362 Seconds
(without alignments)
3648.073 Million cell updates/sec

Title: US-10-028-384-12

Perfect score: 705

Sequence: 1 WTKGFLSLSEKQDTLLKL.....HMLVRIYKVKDLNRLSRT 705

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 824507 seqs, 355394441 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1643622

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10028384/runat 14122004 131646 6816/app query.fasta 1.3740
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=Oligo.rni -MINWATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10028384@cgn 1_1_259@runat 14122004 131646 6816 -NCFU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 128 | 18.2 | 487 | US-09-385-982-213 | Sequence 213, App |
| 2 | 102 | 14.5 | 414 | US-09-513-999C-271 | Sequence 271, App |
| 3 | 102 | 14.5 | 487 | US-09-702-705-1655 | Sequence 1655, App |
| 4 | 102 | 14.5 | 487 | US-09-736-457-1655 | Sequence 1655, App |
| 5 | 102 | 14.5 | 487 | US-09-614-124B-1655 | Sequence 1655, App |
| 6 | 102 | 14.5 | 487 | US-09-671-325-1655 | Sequence 1655, App |
| 7 | 102 | 14.5 | 487 | US-09-658-824-1655 | Sequence 1655, App |
| 8 | 101 | 14.3 | 307 | US-09-702-705-375 | Sequence 375, App |
| 9 | 101 | 14.3 | 307 | US-09-702-705-1271 | Sequence 1271, App |
| 10 | 101 | 14.3 | 307 | US-09-736-457-375 | Sequence 375, App |
| 11 | 101 | 14.3 | 307 | US-09-736-457-1271 | Sequence 1271, App |
| 12 | 101 | 14.3 | 307 | US-09-614-124B-375 | Sequence 375, App |

| | | | | | | |
|----|-----|------|------|---|----------------------|---------------------|
| 13 | 101 | 14.3 | 307 | 4 | US-09-614-124B-1271 | Sequence 1271, App |
| 14 | 101 | 14.3 | 307 | 4 | US-09-671-325-375 | Sequence 375, App |
| 15 | 101 | 14.3 | 307 | 4 | US-09-671-325-1271 | Sequence 1271, App |
| 16 | 101 | 14.3 | 307 | 4 | US-09-589-184-375 | Sequence 375, App |
| 17 | 101 | 14.3 | 307 | 4 | US-09-658-824-375 | Sequence 375, App |
| 18 | 101 | 14.3 | 307 | 4 | US-09-658-824-1271 | Sequence 1271, App |
| 19 | 65 | 9.2 | 307 | 4 | US-09-702-705-588 | Sequence 588, App |
| 20 | 65 | 9.2 | 307 | 4 | US-09-736-457-588 | Sequence 588, App |
| 21 | 65 | 9.2 | 307 | 4 | US-09-614-124B-588 | Sequence 588, App |
| 22 | 65 | 9.2 | 307 | 4 | US-09-671-325-588 | Sequence 588, App |
| 23 | 65 | 9.2 | 307 | 4 | US-09-589-184-588 | Sequence 588, App |
| 24 | 65 | 9.2 | 307 | 4 | US-09-658-824-588 | Sequence 588, App |
| 25 | 39 | 5.5 | 914 | 4 | US-09-270-767-12856 | Sequence 12856, App |
| 26 | 30 | 4.3 | 1386 | 4 | US-09-270-767-11648 | Sequence 11648, App |
| 27 | 26 | 3.7 | 426 | 4 | US-09-513-999C-1213 | Sequence 1213, App |
| 28 | 26 | 3.7 | 503 | 4 | US-09-270-767-27462 | Sequence 27462, App |
| 29 | 21 | 3.0 | 900 | 4 | US-09-270-767-12331 | Sequence 12331, App |
| 30 | 19 | 2.7 | 291 | 4 | US-09-313-294A-4834 | Sequence 4834, App |
| 31 | 17 | 2.4 | 268 | 4 | US-09-313-294A-3259 | Sequence 3259, App |
| 32 | 16 | 2.3 | 2157 | 4 | US-09-614-221A-318 | Sequence 318, App |
| 33 | 15 | 2.1 | 109 | 4 | US-09-270-767-28703 | Sequence 28703, App |
| 34 | 15 | 2.1 | 594 | 4 | US-09-248-796A-3090 | Sequence 3090, App |
| 35 | 13 | 1.8 | 292 | 4 | US-09-313-294A-4442 | Sequence 4442, App |
| 36 | 13 | 1.8 | 305 | 4 | US-09-313-294A-6869 | Sequence 6869, App |
| 37 | 12 | 1.7 | 245 | 4 | US-09-513-999C-21090 | Sequence 21090, App |
| 38 | 12 | 1.7 | 299 | 4 | US-09-313-294A-1303 | Sequence 1303, App |
| 39 | 12 | 1.7 | 507 | 4 | US-09-513-999C-1965 | Sequence 1965, App |
| 40 | 12 | 1.7 | 560 | 4 | US-09-270-767-1318 | Sequence 1318, App |
| 41 | 12 | 1.7 | 560 | 4 | US-09-270-767-16600 | Sequence 16600, App |
| 42 | 12 | 1.7 | 615 | 3 | US-09-328-111-332 | Sequence 332, App |
| 43 | 12 | 1.7 | 867 | 4 | US-09-248-796A-3089 | Sequence 3089, App |
| 44 | 12 | 1.7 | 1660 | 4 | US-09-270-767-158 | Sequence 158, App |
| 45 | 12 | 1.7 | 1660 | 4 | US-09-270-767-158 | Sequence 158, App |

ALIGNMENTS

RESULT 1
US-09-385-982-213
; Sequence 213, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON C., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDNA-260XX
; CURRENT FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER FILING DATE: 1999-06-08
; EARLIER FILING DATE: 1999-01-27
; EARLIER FILING DATE: 1999-01-27
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-385-982-213

Alignment Scores:
Pred. No.: 7,67e-121 Length: 487
Score: 128.00 Matches: 128
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.16% Indels: 0
DB: 3 Gaps: 0

US-10-028-384-12 (1-705) x US-09-385-982-213 (1-487)

QY 271 PheGlyValPheGlyLeuCysGlnIleHisAlaPheValAspTyrLeuArgSerLysLeu 290

Db 88 TTGGGGCTTTGGTCTCTGCTCCAGATCCATGCTTTGGATTACCTGGCGAGCAAGTTG 147
 Qy 291 AsnProGlnGlnPheGluValLeuPheArgSerValIleSerLeuValGlyPheValLeu 310
 Db 148 AATCCACAACTTTGAAGTTCTTTTCCGAGCGCTCATCTCTCTGGTAGGCTTTGTGCTT 207
 Qy 311 LeuThrValGlyAlaLeuLeuMetLeuThrGlyLysIleSerProThrPheArgPhe 330
 Db 208 CTCACCGTGGGAGCTCTCTCATGCTCACAGGAAATATCTCCCTGGACGGGGCTTTC 267
 Qy 331 TyrSerLeuLeuAspProSerTyrAlaLysAsnIleProIleIleIleAsnValSer 350
 Db 268 TACTCAGTGCCTGGATGCCCTCTATGCTTAAGAACAAATCCCATCTCTCTGTGCT 327
 Qy 351 GluHisGlnProThrThrTpsSerTyrTyrPheAspLeuGlnLeuValPheMet 370
 Db 328 GAGCATCAGCCCAACCTGGTCTCTCATCTATTTTGACCTGCAGCTCTCTGCTCATG 387
 Qy 371 PheProValGlyLeuTyrTyrCysPheSerAsnLeuSerAspAlaArgIlePheIleIle 390
 Db 388 TTTCAGTGGCTCTATTAAGTCTTTAGCAACCTCTGTGATGCCCGGATTTTATCATC 447
 Qy 391 MetTyrGlyValThrSerMetTyr 398
 Db 448 ATGTATGGTGTGACCATGATG 471

RESULT 2

US-09-513-999C-271
 ; Sequence 271, Application US/09513999C
 ; Patent No. 6783961
 ; GENERAL INFORMATION:
 ; APPLICANT: Dumas Milne Edwards, J.B.
 ; APPLICANT: Duclert, A.
 ; APPLICANT: Giordano, J.Y.
 ; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
 ; FILE REFERENCE: 59.US2.REG
 ; CURRENT APPLICATION NUMBER: US/09/513,999C
 ; CURRENT FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/122,487
 ; PRIOR FILING DATE: 1999-02-26
 ; NUMBER OF SEQ ID NOS: 36681
 ; SOFTWARE: Patent.pm
 ; SEQ ID NO 271
 ; LENGTH: 414
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 108..413
 ; FEATURE:
 ; NAME/KEY: sig_peptide
 ; LOCATION: 108..197
 ; OTHER INFORMATION: score 7.3
 ; OTHER INFORMATION: seq LKLLLSMAAVLS/FS

US-09-513-999C-271
 Alignment Scores:
 Pred. No.: 1,81e-94 Length: 414
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-12 (1-705) x US-09-513-999C-271 (1-414)
 Qy 1 MetThrLysPheGlyPheLeuArgLeuSerTyrGlyLysGlnAspThrLeuLeuLysLeu 20
 Db 108 ATGACTAAGTTTGGATTGTTTGGCATTTGCTATGAGAACGAGACACATTTTGAAGCTT 167
 Qy 21 LeuIleLeuSerMetAlaAlaValLeuSerPheSerThrArgLeuPheAlaValLeuArg 40

Db 168 CTCATTCGTGCAATGGCTGCTGCTATTATTCCTTCTCCACTGCTGCTTTCCTGCTGAGA 227
 Qy 41 PheGluSerValIleHisGluPheAspProTyrPheAsnTyrArgThrThrArgPheLeu 60
 Db 228 TTTGAAGTGTATCATCATGAGTTTCATCGTACTTAAATTAATTCGAGCTACCAAGTTTCTGT 287
 Qy 61 AlaGluGluGlyPheTyrLysPheHisAsnTyrPheAspAspAlaTyrTyrProLeu 80
 Db 288 GCTGAGAGGGGTTTATAAATTCATAACTGGTTTGTATGACCGAGCTCGTACCTTTG 347
 Qy 81 GlyArgIleGlyGlyThrIleTyrProGlyLeuMetIleThrSerAlaAlaIleTyr 100
 Db 348 GGACGAATCATTTGGAGGAACAATTTACCCAGGTTTAATGATCACCTCTGCTGCAATCTAC 407
 Qy 101 HisVal 102
 Db 408 CATGTA 413

RESULT 3

US-09-702-705-1655
 ; Sequence 1655, Application US/09702705
 ; Patent No. 6504010
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Fan, Liqun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCES: 210121.478C14
 ; CURRENT APPLICATION NUMBER: US/09/702,705
 ; CURRENT FILING DATE: 2000-10-30
 ; NUMBER OF SEQ ID NOS: 1833
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1655
 ; LENGTH: 487
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-702-705-1655

Alignment Scores:
 Pred. No.: 2.12e-94 Length: 487
 Score: 102.00 Matches: 102
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 14.47% Indels: 0
 DB: 4 Gaps: 0
 US-10-028-384-12 (1-705) x US-09-702-705-1655 (1-487)
 Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIlePhe 502
 Db 77 TACTCTTCTCCGTCCATTTGCTATCTCCCGTGGTGGGATGGCAGTAGATCATATTT 136
 Qy 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal 522
 Db 137 GATGACTTCCGAGAAGCATATTTATTTGGTCCGTCATATATCTCCAGAGGATGCCAAGTC 196
 Qy 523 MetSerTyrTyrAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
 Db 197 ATGTCTCTGGTGGATTAAGGCTATCAGATTACAGTATGGCAACCCGACAACTTTTAGTG 256
 Qy 543 AspAsnAsnThrThrPheAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
 Db 257 GACATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAGGCAATGGCGTCCACA 316
 Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582

Db 317 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTGCTGATGTGCTGCTATTTT 376
Qy 583 GlyGly 584
Db 377 GGAGGA 382

RESULT 4

US-09-736-457-1655
; Sequence 1655, Application US/09736457
; Patent No. 6509448

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriek
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736.457

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 1864

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1655

; LENGTH: 487

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-736-457-1655

Alignment Scores:

Pred. No.: 2,12e-94 Length: 487
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.47% Indels: 0
DB: Gaps: 0

US-10-028-384-12 (1-705) x US-09-736-457-1655 (1-487)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
Db 77 TACTCTTCTCGTCCATGTACTATCTGCCCGTGGTGGGATGGCAGTAGGATCATATTT 136
Qy 503 AspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 137 GATGACTTCGAGAGCATATATTGGCTCCGTCAATAATCTCCAGAGGATCGAAGGTC 196

Qy 523 MetSerTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 197 ATGTCCTGGTGGATATGGCTATCAGATTACAGTATGCAACCCGCAATTTTAGTG 256

Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 257 GACAATAACACATGGAATAATACCATATTTCTCGAGTAGGAGCAGCATGGCTCCACA 316

Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 317 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTGCTGCTATGCTGCTATTTT 376

Qy 583 GlyGly 584

Db 377 GGAGGA 382

RESULT 5

US-09-614-124B-1655

; Sequence 1655, Application US/09614124B

; Patent No. 6630574

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedvick, Tom

; APPLICANT: Carter, Darriek

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C9

; CURRENT APPLICATION NUMBER: US/09/614,124B

; CURRENT FILING DATE: 2001-07-11

; NUMBER OF SEQ ID NOS: 1668

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1655

; LENGTH: 487

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-614-124B-1655

Alignment Scores:

Pred. No.: 2,12e-94 Length: 487
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.47% Indels: 0
DB: Gaps: 0

US-10-028-384-12 (1-705) x US-09-614-124B-1655 (1-487)

Qy 483 TyrSerSerProSerIleValLeuSerAlaArgGlyAspGlySerArgIleIlePhe 502
Db 77 TACTCTTCTCGTCCATGTACTATCTGCCCGTGGTGGGATGGCAGTAGGATCATATTT 136

Qy 503 AspAspPheArgGluAlaTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522

Db 137 GATGACTTCGAGAGCATATATTGGCTCCGTCAATAATCTCCAGAGGATCGAAGGTC 196

Qy 523 MetSerTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542

Db 197 ATGTCCTGGTGGATATGGCTATCAGATTACAGTATGCAACCCGCAATTTTAGTG 256

Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562

Db 257 GACAATAACACATGGAATAATACCATATTTCTCGAGTAGGAGCAGCAATGGCTCCACA 316

Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582

Db 317 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTGCTGCTATGCTGCTATTTT 376

Qy 583 GlyGly 584

Db 377 GGAGGA 382

RESULT 6

US-09-671-325-1655

; Sequence 1655, Application US/09671325

; Patent No. 6867154

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedvick, Tom

; APPLICANT: Carter, Darriek

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C12

487

QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
DB 63 GATGACTTCGAGAGCATATTATGGCTCCGTCTAATACTCCAGAGATCGGAGGTC 122
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
DB 123 ATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTAGTG 182
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
DB 183 GACATAACACATGGAATAATACCCATATTTCTCAGTAGGCGAGCAATGGCGTCCACA 242
QY 563 GluGluLysAlaTyrGluIleValMetArgLuleuAspValSerTyrValLeuValIlePhe 582
DB 243 GAGGAAAAAGCCTATGAGATCATGAGGAGGCTCGATGTCAGCTATGCTGGTCAATTTT 302
QY 583 Gly 583
DB 303 GGA 305

RESULT 9

US-09-702-705-1271
; Sequence 1271, Application US/09702705
; Patent No. 6504010

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14

; CURRENT APPLICATION NUMBER: US/09/702,705

; CURRENT FILING DATE: 2000-10-30

; NUMBER OF SEQ ID NOS: 1833

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1271

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-702-705-1271

Alignment Scores:

Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-702-705-1271 (1-307)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
DB 3 TACTCTTCCTCCGTCATATGCTATCTGCGCGTGGTGGGATGCGAGTAGGATCATATTT 62
QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
DB 63 GATGACTTCGAGAGCATATTATGGCTTCGTCTAATACTCCAGAGATCGGAGGTC 122
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
DB 123 ATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTAGTG 182
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
DB 183 GACATAACACATGGAATAATACCCATATTTCTCAGTAGGCGAGCAATGGCGTCCACA 242

RESULT 11

US-09-736-457-1271

; Sequence 1271, Application US/09736457

QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
DB 243 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTT 302
QY 583 Gly 583
DB 303 GGA 305

RESULT 10

US-09-736-457-375

; Sequence 375, Application US/09736457

; Patent No. 6509448

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15

; CURRENT APPLICATION NUMBER: US/09/736,457

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 1864

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 375

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-736-457-375

Alignment Scores:

Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-736-457-375 (1-307)

QY 483 TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
DB 3 TACTCTTCCTCCGTCATATGCTATCTGCGCGTGGTGGGATGCGAGTAGGATCATATTT 62
QY 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
DB 63 GATGACTTCGAGAGCATATTATTTGGCTCGTCTAATACTCCAGAGATCGGAGGTC 122
QY 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
DB 123 ATGTCCTGGTGGGATTATGGCTATCAGATTACAGCTATGCGCAACCGAACAAATTTAGTG 182
QY 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
DB 183 GACATAACACATGGAATAATACCCATATTTCTCGAGTAGGCGAGCAATGGCGTCCACA 242
QY 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
DB 243 GAGGAAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTT 302
QY 583 Gly 583
DB 303 GGA 305


```

; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriek
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-736-457-1271

Alignment Scores:
Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-736-457-1271 (1-307)

Qy 483 TyrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCGTCATTTGCTATCTGCGCGTGGGGATGGCGATAGGATCATATTT 62
Qy 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 63 GATGACTTCGAGAGCATATTTATGGCTCCGTCATATAATACCTACCTCAGAGATGGAGGTC 122
Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 123 ATGCTCTGGTGGGATTTATGGCTATCAGATTACAGTATGGCAACCGAACAATTTAGTG 182
Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACAATAACACATGGGAATAATACCCATATTTCTCGAGTAGGCGAGCAATGGGTCACCA 242
Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTT 302
Qy 583 Gly 583
Db 303 GGA 305

RESULT 12
US-09-614-124B-375
; Sequence 375, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriek
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-375

Alignment Scores:

```

```

; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 375
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-375

Alignment Scores:
Pred. No.: 1.42e-93 Length: 307
Score: 101.00 Matches: 101
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.33% Indels: 0
DB: 4 Gaps: 0

US-10-028-384-12 (1-705) x US-09-614-124B-375 (1-307)

Qy 483 TyrSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe 502
Db 3 TACTCTTCCTCGTCATTTGCTATCTGCGCGTGGGGATGGCGATAGGATCATATTT 62
Qy 503 AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaVal 522
Db 63 GATGACTTCGAGAGCATATTTATGGCTCCGTCATATAATACCTACCTCAGAGATGGAGGTC 122
Qy 523 MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal 542
Db 123 ATGCTCTGGTGGGATTTATGGCTATCAGATTACAGTATGGCAACCGAACAATTTAGTG 182
Qy 543 AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr 562
Db 183 GACAATAACACATGGGAATAATACCCATATTTCTCGAGTAGGCGAGCAATGGGTCACCA 242
Qy 563 GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe 582
Db 243 GAGGAAAAGCCTATGAGATCATGAGGAGCTCGATGTCAGCTATGCTGGTCAATTTT 302
Qy 583 Gly 583
Db 303 GGA 305

RESULT 13
US-09-614-124B-1271
; Sequence 1271, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darriek
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1271
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-614-124B-1271

Alignment Scores:

```

| | | | |
|----|-----|--|-----|
| Qy | 563 | GluGluLysAlaTyrGluIleMetArgGluLeuAspValSerTyrValLeuValIlePhe | 582 |
| Db | 243 | GAGGAAAAGCCTAAGAGATCATGAGGAGCTCGATGTCAGCTATGCTGCTCATTTT | 302 |
| Qy | 583 | Gly | 583 |
| Db | 303 | GGA | 305 |

Search completed: December 14, 2004, 23:43:59
 Job time : 141.362 secs

| | | | | | |
|-----|----|--|-----|----|---|
| 483 | Qy | TyrSerSerProSerIleValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe | 502 | Qy | ValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe |
| 3 | Db | TACTCTTCCTCCGTCCTACTATCTCCCGCTGGTGGGATGGCAGTAGCATATTT | 62 | Db | TACTCTTCCTCCGTCCTACTATCTCCCGCTGGTGGGATGGCAGTAGCATATTT |
| 503 | Qy | AspAspPheArgGluAlaTyrTyrTrpLeuArgHisAsnThrProGluAspAlaLysVal | 522 | Qy | ValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe |
| 63 | Db | GATGACTCTCCGAGAGCATATTATTGGCTTCTCATATACTCCAGAGGATCGAAGGTC | 122 | Db | TACTCTTCCTCCGTCCTACTATCTCCCGCTGGTGGGATGGCAGTAGCATATTT |
| 523 | Qy | MetSerTrpTrpAspTyrGlyTyrGlnIleThrAlaMetAlaAsnArgThrIleLeuVal | 542 | Qy | ValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe |
| 123 | Db | ATGCTCTGGTGGGATTAATGGCTATCAGATTACAGCTATGGCAACCAATTTAGTG | 182 | Db | TACTCTTCCTCCGTCCTACTATCTCCCGCTGGTGGGATGGCAGTAGCATATTT |
| 543 | Qy | AspAsnAsnThrTrpAsnAsnThrHisIleSerArgValGlyGlnAlaMetAlaSerThr | 562 | Qy | ValLeuSerAlaArgGlyGlyAspGlySerArgIleIlePhe |
| 183 | Db | GACAAATAACACATGGAATTAATACCATATTTCTCGAGTAGGCGAGGCAATGGCGTCACA | 242 | Db | TACTCTTCCTCCGTCCTACTATCTCCCGCTGGTGGGATGGCAGTAGCATATTT |